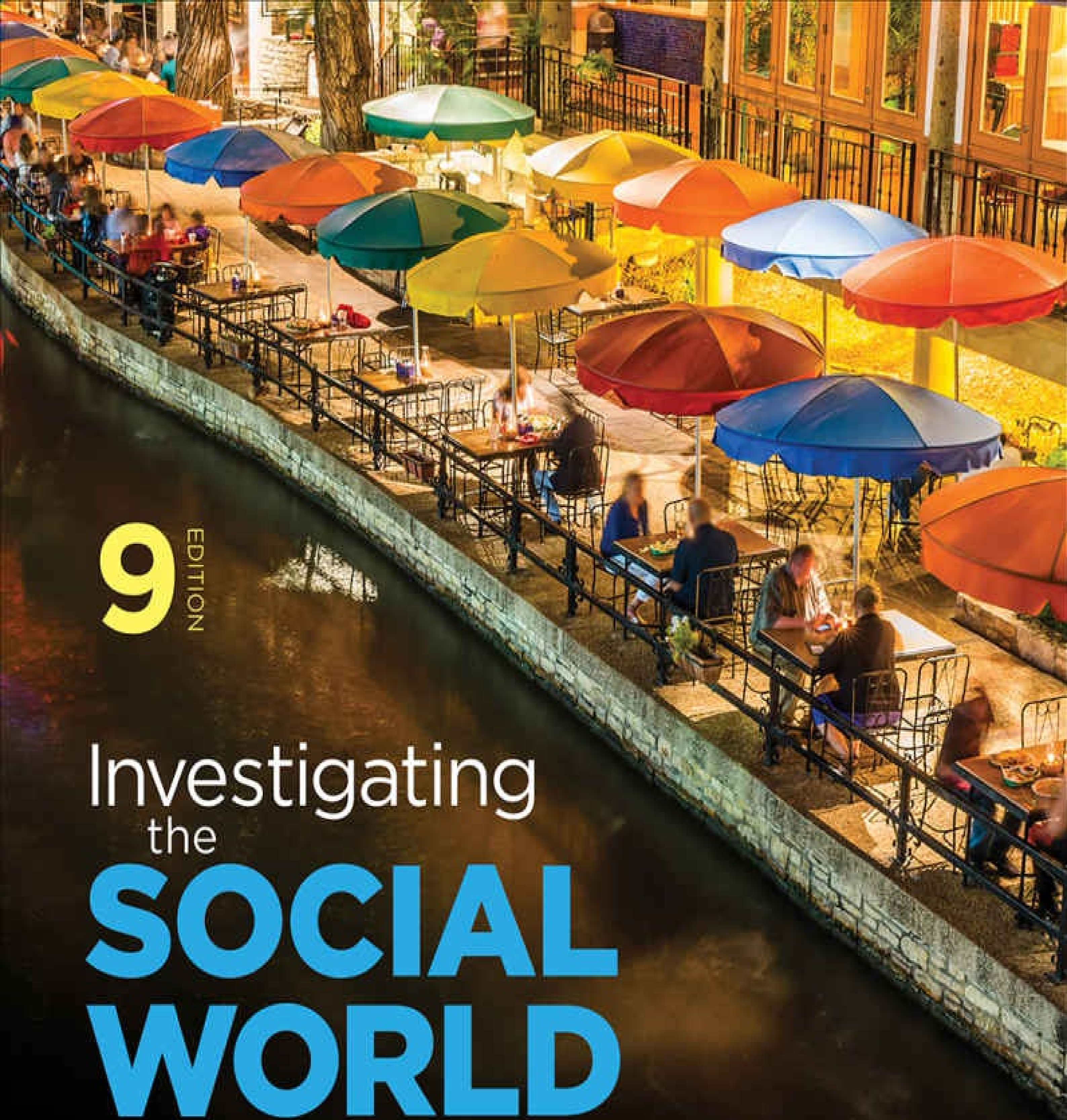


9 EDITION

Investigating the SOCIAL WORLD

The Process and Practice of Research

Russell K. Schutt

Investigating the Social World

Ninth Edition

To Julia Ellen Schutt

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Investigating the Social World

The Process and Practice of Research

Ninth Edition

Russell K. Schutt
University of Massachusetts Boston

SAGE
Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.

2455 Teller Road

Thousand Oaks, California 91320

E-mail: order@sagepub.com

SAGE Publications Ltd.

1 Oliver's Yard

55 City Road

London EC1Y 1SP

United Kingdom

SAGE Publications India Pvt. Ltd.

B 1/I 1 Mohan Cooperative Industrial Area

Mathura Road, New Delhi 110 044

India

SAGE Publications Asia-Pacific Pte. Ltd.

3 Church Street

#10–04 Samsung Hub

Singapore 049483

Copyright © 2019 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All trademarks depicted within this book, including trademarks appearing as part of a screenshot, figure, or other image are included solely for the purpose of illustration and are the property of their respective holders. The use of the trademarks in no way indicates any relationship with, or endorsement by, the holders of said trademark.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Schutt, Russell K., author.

Title: Investigating the social world : the process and practice of research / Russell K. Schutt, University of Massachusetts Boston.

Description: Ninth Edition. | Thousand Oaks : SAGE Publications, [2018] | Revised edition of the author's Investigating the social world, [2015] | Includes bibliographical references and indexes.

Identifiers: LCCN 2017060167 | ISBN 9781506361192 (pbk. : alk. paper)

Subjects: LCSH: Social problems—Research. | Social sciences—Research.

Classification: LCC HN29 .S34 2018 | DDC 361.1072—dc23 LC record available at https://lccn.loc.gov/2017060167

This book is printed on acid-free paper.

18 19 20 21 22 10 9 8 7 6 5 4 3 2 1

Publisher: Jeff Lasser

Assistant Content Development Editor: Sarah Dillard

Editorial Assistant: Adeline Wilson

Marketing Manager: Kara Kindstrom

Production Editor: Veronica Stapleton Hooper

Copy Editor: Amy Marks

Typesetter: C&M Digitals (P) Ltd.

Proofreader: Dennis W. Webb

Indexer: Sheila Bodell

Cover Designer: Candice Harman

Brief Contents

1. About the Author
2. Preface
3. Acknowledgments
4. Section I. Foundations for Social Research
    1. 1. Science, Society, and Social Research
    2. 2. The Process and Problems of Social Research
    3. 3. Research Ethics and Research Proposals
5. Section II. Fundamentals of Social Research
    1. 4. Conceptualization and Measurement
    2. 5. Sampling and Generalizability
    3. 6. Research Design and Causation
6. Section III. Basic Social Research Designs
    1. 7. Experiments
    2. 8. Survey Research
    3. 9. Quantitative Data Analysis
    4. 10. Qualitative Methods
    5. 11. Qualitative Data Analysis
7. Section IV. Complex Social Research Designs
    1. 12. Mixed Methods
    2. 13. Evaluation and Policy Research
    3. 14. Research Using Secondary Data and “Big” Data
    4. 15. Research Using Historical and Comparative Data and Content Analysis
    5. 16. Summarizing and Reporting Research
8. Appendix A: Questions to Ask About a Research Article
9. Appendix B: How to Read a Research Article
10. Appendix C: Table of Random Numbers
11. Glossary
12. Bibliography
13. Index

Detailed Contents

- About the Author
- Preface
- Acknowledgments
- Section I. Foundations for Social Research
  - 1. Science, Society, and Social Research
    - Research That Matters, Questions That Count
    - Learning About the Social World
    - Avoiding Errors in Reasoning About the Social World
      - Observing
      - Generalizing
      - Reasoning
      - Reevaluating
    - Science and Social Science
      - The Scientific Approach
    - Research in the News: Social Media and Political Polarization
      - Pseudoscience or Science
      - Motives for Social Research
      - Types of Social Research
        - Descriptive Research
        - Exploratory Research
        - Explanatory Research
        - Evaluation Research
    - Careers and Research
    - Strengths and Limitations of Social Research
    - Alternative Research Orientations
      - Quantitative and/or Qualitative Methods
      - Philosophical Perspectives
      - Basic Science or Applied Research
      - The Role of Values
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises
    - ■ Ethics Questions
    - ■ Web Exercises
    - ■ Video Interview Questions
    - ■ SPSS Exercises
    - ■ Developing a Research Proposal

- 2. The Process and Problems of Social Research
  - Research That Matters, Questions That Count
  - Social Research Questions
    - Identifying Social Research Questions
    - Refining Social Research Questions
    - Evaluating Social Research Questions
      - Feasibility
      - Social Importance
      - Scientific Relevance
  - Social Theories
    - Scientific Paradigms
  - Social Research Foundations
    - Searching the Literature
    - Reviewing Research
      - Single-Article Reviews: Formal and Informal Deterrents to Domestic Violence
      - Integrated Literature Reviews: When Does Arrest Matter?
      - Systematic Literature Reviews: Second Responder Programs and Repeat Family Abuse Incidents
    - Searching the Web
  - Social Research Strategies
  - Research in the News: Control and Fear: What Mass Killings and Domestic Violence Have in Common
    - Explanatory Research
      - Deductive Research
      - Domestic Violence and the Research Circle
      - Inductive Research
    - Exploratory Research
      - Battered Women's Help Seeking
      - Descriptive Research
  - Careers and Research
  - Social Research Organizations
  - Social Research Standards
    - Measurement Validity
    - Generalizability
    - Causal Validity
    - Authenticity
  - Conclusions
  - ■ Key Terms
  - ■ Highlights
  - ■ Discussion Questions
  - ■ Practice Exercises

- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises
- ■ Developing a Research Proposal

3. Research Ethics and Research Proposals
  - Research That Matters, Questions That Count
  - Historical Background
  - Ethical Principles
    - Achievement of Valid Results
    - Honesty and Openness
    - Protection of Research Participants
      - Avoid Harming Research Participants
      - Obtain Informed Consent
      - Avoid Deception in Research, Except in Limited Circumstances
      - Maintain Privacy and Confidentiality
    - Consider Uses of Research So That Benefits Outweigh Risks
  - The Institutional Review Board
  - Research in the News: Some Social Scientists Are Tired of Asking for Permission
  - Careers and Research
  - Social Research Proposals
    - Case Study: Evaluating a Public Health Program
  - Conclusions
  - ■ Key Terms
  - ■ Highlights
  - ■ Discussion Questions
  - ■ Practice Exercises
  - ■ Ethics Questions
  - ■ Web Exercises
  - ■ Video Interview Questions
  - ■ SPSS Exercises
  - ■ Developing a Research Proposal

Section II. Fundamentals of Social Research

4. Conceptualization and Measurement
  - Research That Matters, Questions That Count
  - Concepts
    - Conceptualization in Practice
      - Substance Abuse
      - Youth Gangs
      - Poverty

- From Concepts to Indicators
- Research in the News: Are Teenagers Replacing Drugs With Smartphones?
  - Abstract and Concrete Concepts
    - Operationalizing the Concept of Race
    - Operationalizing Social Network Position
- From Observations to Concepts
- Measurement
  - Constructing Questions
  - Making Observations
  - Collecting Unobtrusive Measures
  - Using Available Data
  - Coding Content
  - Taking Pictures
  - Combining Measurement Operations
- Careers and Research
- Levels of Measurement
  - Nominal Level of Measurement
  - Ordinal Level of Measurement
  - Interval Level of Measurement
  - Ratio Level of Measurement
  - The Special Case of Dichotomies
  - Comparison of Levels of Measurement
- Evaluating Measures
  - Measurement Validity
    - Face Validity
    - Content Validity
    - Criterion Validity
    - Construct Validity
  - Measurement Reliability
    - Multiple Times: Test–Retest and Alternate Forms
    - Multiple Indicators: Interitem and Split-Half
    - Multiple Observers: Interobserver and Intercoder
  - Ways to Improve Reliability and Validity
- Conclusions
- ■ Key Terms
- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions

- ■ SPSS Exercises
- ■ Developing a Research Proposal

5. Sampling and Generalizability
  - Research That Matters, Questions That Count
  - Sample Planning
    - The Purpose of Sampling
    - Define Sample Components and the Population
    - Evaluate Generalizability
    - Assess the Diversity of the Population
  - Research in the News: What Are Best Practices for Sampling Vulnerable Populations?
    - Consider a Census
  - Sampling Methods
    - Probability Sampling Methods
      - Simple Random Sampling
      - Systematic Random Sampling
      - Stratified Random Sampling
      - Multistage Cluster Sampling
      - Probability Sampling Methods Compared
    - Nonprobability Sampling Methods
      - Availability (Convenience) Sampling
  - Careers and Research
    - Quota Sampling
    - Purposive Sampling
    - Snowball Sampling
    - Lessons About Sample Quality
    - Generalizability in Qualitative Research
  - Sampling Distributions
    - Estimating Sampling Error
    - Sample Size Considerations
  - Conclusions
  - ■ Key Terms
  - ■ Highlights
  - ■ Discussion Questions
  - ■ Practice Exercises
  - ■ Ethics Questions
  - ■ Web Exercises
  - ■ Video Interview Questions
  - ■ SPSS Exercises
  - ■ Developing a Research Proposal

6. Research Design and Causation
  - Research That Matters, Questions That Count

- Research Design Alternatives
  - Units of Analysis
    - Individual and Group
    - The Ecological Fallacy and Reductionism
- Research in the News: Police and Black Drivers
  - Cross-Sectional and Longitudinal Designs
    - Cross-Sectional Designs
    - Longitudinal Designs
  - Quantitative or Qualitative Causal Explanations
    - Quantitative (Nomothetic) Causal Explanations
    - Qualitative (Idiographic) Causal Explanations
- Careers and Research
- Criteria and Cautions for Nomothetic Causal Explanations
  - Association
  - Time Order
    - Experimental Designs
    - Nonexperimental Designs
  - Nonspuriousness
    - Randomization
    - Statistical Control
  - Mechanism
  - Context
- Comparing Research Designs
- Conclusions
- ■ Key Terms
- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises
- ■ Developing a Research Proposal

Section III. Basic Social Research Design

- 7. Experiments
  - Research That Matters, Questions That Count
  - History of Experimentation
  - Careers and Research
  - True Experiments
    - Experimental and Comparison Groups
    - Pretest and Posttest Measures
    - Randomization

        - Limitations of True Experimental Designs
        - Summary: Causality in True Experiments
    - Quasi-Experiments
        - Nonequivalent Control Group Designs
    - Research in the News: Airbnb Hosts and the Disabled
        - Aggregate Matching
        - Individual Matching
        - Ex Post Facto Control Group Designs
        - Before-and-After Designs
        - Summary: Causality in Quasi-Experiments
    - Validity in Experiments
        - Causal (Internal) Validity
            - Sources of Internal Invalidity Reduced by a Comparison Group
            - Sources of Internal Invalidity Reduced by Randomization
            - Sources of Internal Invalidity That Require Attention While the Experiment Is in Progress
        - Generalizability
            - Sample Generalizability
            - Factorial Surveys
            - External Validity
            - Interaction of Testing and Treatment
    - Ethical Issues in Experimental Research
        - Deception
        - Selective Distribution of Benefits
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises
    - ■ Ethics Questions
    - ■ Web Exercises
    - ■ Video Interview Questions
    - ■ SPSS Exercises
    - ■ Developing a Research Proposal
- 8. Survey Research
    - Research That Matters, Questions That Count
    - Survey Research in the Social Sciences
        - Attractions of Survey Research
            - Versatility
            - Efficiency
            - Generalizability

    - The Omnibus Survey
    - Errors in Survey Research
- Writing Survey Questions
    - Avoid Confusing Phrasing
    - Minimize the Risk of Bias
    - Maximize the Utility of Response Categories
    - Avoid Making Either Disagreement or Agreement Disagreeable
    - Minimize Fence-Sitting and Floating
- Combining Questions in Indexes
- Designing Questionnaires
    - Build on Existing Instruments
    - Refine and Test Questions
    - Add Interpretive Questions
- Careers and Research
    - Maintain Consistent Focus
- Research in the News: Social Interaction Critical for Mental and Physical Health
    - Order the Questions
    - Make the Questionnaire Attractive
    - Consider Translation
- Organizing Surveys
    - Mailed, Self-Administered Surveys
    - Group-Administered Surveys
    - Telephone Surveys
        - Reaching Sample Units
        - Maximizing Response to Phone Surveys
    - In-Person Interviews
        - Balancing Rapport and Control
        - Maximizing Response to Interviews
    - Web Surveys
    - Mixed-Mode Surveys
    - A Comparison of Survey Designs
- Ethical Issues in Survey Research
- Conclusions
- ■ Key Terms
- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises

        - ■ Developing a Research Proposal
- 9. Quantitative Data Analysis
    - Research That Matters, Questions That Count
    - Introducing Statistics
        - Case Study: The Likelihood of Voting
    - Preparing for Data Analysis
    - Displaying Univariate Distributions
        - Graphs
        - Frequency Distributions
            - Ungrouped Data
            - Grouped Data
            - Combined and Compressed Distributions
    - Summarizing Univariate Distributions
    - Research in the News: Why Key State Polls Were Wrong About Trump
        - Measures of Central Tendency
            - Mode
            - Median
            - Mean
            - Median or Mean?
        - Measures of Variation
            - Range
            - Interquartile Range
            - Variance
            - Standard Deviation
    - Analyzing Data Ethically: How Not to Lie With Statistics
    - Cross-Tabulating Variables
        - Constructing Contingency Tables
        - Graphing Association
        - Describing Association
        - Evaluating Association
        - Controlling for a Third Variable
            - Intervening Variables
            - Extraneous Variables
            - Specification
    - Careers and Research
    - Regression Analysis
    - Performing Meta-Analyses
        - Case Study: Patient–Provider Race Concordance and Minority Health Outcomes
    - Analyzing Data Ethically: How Not to Lie About Relationships
    - Conclusions
    - ■ Key Terms

- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises
- ■ Developing a Research Proposal

10. Qualitative Methods
    - Research That Matters, Questions That Count
    - Fundamentals of Qualitative Methods
        - History of Qualitative Research
        - Features of Qualitative Research
        - Basics of Qualitative Research
            - The Case Study
            - Ethnography
    - Careers and Research
        - Digital Ethnography
    - Participant Observation
        - Choosing a Role
            - Covert Observation
            - Overt Observation
            - Overt Participation (Participant Observer)
            - Covert Participation
    - Research in the News: Family Life on Hold After Hurricane Harvey
        - Entering the Field
        - Developing and Maintaining Relationships
        - Sampling People and Events
        - Taking Notes
        - Managing the Personal Dimensions
    - Intensive Interviewing
        - Establishing and Maintaining a Partnership
        - Asking Questions and Recording Answers
    - Interviewing Online
    - Focus Groups
    - Generalizability in Qualitative Research
    - Ethical Issues in Qualitative Research
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises

- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises
- ■ Developing a Research Proposal

11. Qualitative Data Analysis
  - Research That Matters, Questions That Count
  - Features of Qualitative Data Analysis
    - Qualitative Data Analysis as an Art
    - Qualitative Compared With Quantitative Data Analysis
  - Techniques of Qualitative Data Analysis
    - Documentation
    - Organization, Categorization, and Condensation
    - Examination and Display of Relationships
    - Corroboration and Legitimization of Conclusions
    - Reflection on the Researcher's Role
  - Alternatives in Qualitative Data Analysis
    - Grounded Theory
    - Abductive Analysis
    - Case-Oriented Understanding
  - Research in the News: How to Understand Solitary Confinement
    - Conversation Analysis
    - Narrative Analysis
    - Ethnomethodology
    - Qualitative Comparative Analysis
    - Combining Qualitative Methods
  - Visual Sociology
  - Careers and Research
  - Systematic Observation
  - Participatory Action Research
  - Computer-Assisted Qualitative Data Analysis
  - Ethics in Qualitative Data Analysis
  - Conclusions
  - ■ Key Terms
  - ■ Highlights
  - ■ Discussion Questions
  - ■ Practice Exercises
  - ■ Ethics Questions
  - ■ Web Exercises
  - ■ Video Interview Questions
  - ■ HyperRESEARCH Exercises
  - ■ Developing a Research Proposal

- Section IV. Complex Social Research Designs
  - 12. Mixed Methods
    - Research That Matters, Questions That Count
    - History of Mixed Methods
    - Types of Mixed Methods
      - Integrated Mixed-Methods Designs
      - Embedded Mixed-Methods Designs
      - Staged Mixed-Methods Designs
      - Complex Mixed-Methods Designs
    - Strengths and Limitations of Mixed Methods
    - Research in the News: Why Women Don’t Report Sexual Harassment
    - Careers and Research
    - Ethics and Mixed Methods
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises
    - ■ Ethics Questions
    - ■ Web Exercises
    - ■ Video Interview Questions
    - ■ SPSS Exercises
    - ■ Developing a Research Proposal
  - 13. Evaluation and Policy Research
    - Research That Matters, Questions That Count
    - History of Evaluation Research
    - Evaluation Basics
    - Questions for Evaluation Research
      - Needs Assessment
    - Research in the News: No-Cost Talk Therapy?
      - Evaluability Assessment
      - Process Evaluation
      - Impact Analysis
      - Efficiency Analysis
    - Design Decisions
      - Black Box Evaluation or Program Theory
    - Careers and Research
      - Researcher or Stakeholder Orientation
      - Quantitative or Qualitative Methods
      - Simple or Complex Outcomes
      - Groups or Individuals
    - Policy Research

- Ethics in Evaluation
- Conclusions
- ■ Key Terms
- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercises
- ■ Developing a Research Proposal

14. Research Using Secondary Data and “Big” Data
- Secondary Data Sources
- Research That Matters, Questions That Count
- Careers and Research
  - U.S. Census Bureau
  - Integrated Public Use Microdata Series
  - Bureau of Labor Statistics
  - Other Government Sources
  - Other Data Sources
  - Inter-university Consortium for Political and Social Research
    - Types of Data Available From ICPSR
    - Obtaining Data From ICPSR
  - Harvard’s Dataverse
  - International Data Sources
  - Qualitative Data Sources
- Challenges for Secondary Data Analyses
- Big Data
  - Background
  - Examples of Research Using Big Data
- Ethical Issues in Secondary Data Analysis and Big Data
- Research in the News: A Bright Side to Facebook’s Experiments on Its Users?
- Conclusions
- ■ Key Terms
- ■ Highlights
- ■ Discussion Questions
- ■ Practice Exercises
- ■ Ethics Questions
- ■ Web Exercises
- ■ Video Interview Questions
- ■ SPSS Exercise

    - ■ Developing a Research Proposal
- 15. Research Using Historical and Comparative Data and Content Analysis
    - Research That Matters, Questions That Count
    - Overview of Historical and Comparative Research Methods
    - Historical Social Science Methods
        - Historical Events Research
            - Event-Structure Analysis
            - Oral History
        - Historical Process Research
        - Cautions for Historical Methods
    - Comparative Social Science Methods
    - Research in the News: Britain Cracking Down on Gender Stereotypes in Ads
        - Cross-Sectional Comparative Research
    - Careers and Research
        - Comparative Historical Research
        - Comparative Case Study Designs
        - Cautions for Comparative Methods
    - Demographic Analysis
    - Content Analysis
        - Identify a Population of Documents or Other Textual Sources
        - Determine the Units of Analysis
        - Select a Sample of Units From the Population
        - Design Coding Procedures for the Variables to be Measured
        - Develop Appropriate Statistical Analyses
    - Ethical Issues in Historical and Comparative Research and Content Analysis
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises
    - ■ Ethics Questions
    - ■ Web Exercises
    - ■ Video Interview Questions
    - ■ SPSS Exercises
    - ■ Developing a Research Proposal
- 16. Summarizing and Reporting Research
    - Research That Matters, Questions That Count
    - Writing Research
    - Displaying Research
    - Reporting Research

        - Journal Articles
    - Research in the News: Do Preschool Teachers Need to Be College Graduates?
        - Applied Research Reports
            - Findings From California's Welcome Baby Program
            - Limitations
            - Conclusions
            - Framing an Applied Report
        - Research Posters
        - Reporting Quantitative and Qualitative Research
    - Ethics, Politics, and Research Reports
    - Careers and Research
        - Communicating With the Public
        - Plagiarism
    - Conclusions
    - ■ Key Terms
    - ■ Highlights
    - ■ Discussion Questions
    - ■ Practice Exercises
    - ■ Ethics Questions
    - ■ Web Exercises
    - ■ Video Interview Questions
    - ■ SPSS Exercises
    - ■ Developing a Research Proposal
- Appendix A: Questions to Ask About a Research Article
- Appendix B: How to Read a Research Article
- Appendix C: Table of Random Numbers
- Glossary
- Bibliography
- Index

About the Author

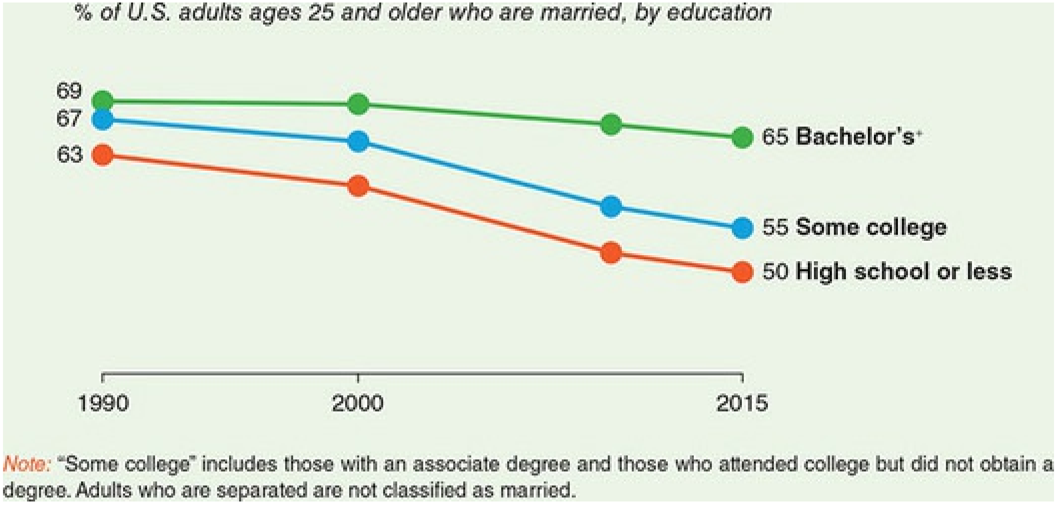
Russell K. Schutt,

PhD, is Professor of Sociology at the University of Massachusetts Boston; Research Associate in Psychiatry at the Beth Israel Deaconess Medical Center, Harvard Medical School; and Research Associate at Edith Nourse Rogers Memorial Veterans Hospital, Department of Veterans Affairs. He completed his BA, MA, and PhD degrees at the University of Illinois at Chicago and his postdoctoral fellowship in the Sociology of Social Control Training Program at Yale University. In addition to *Investigating the Social World: The Process and Practice of Research* and adaptations of that text—*Understanding the Social World: Research Methods for the 21st Century, Making Sense of the Social World* (with Dan Chambliss), *Research Methods in Psychology* (with Paul G. Nestor), *The Practice of Research in Criminology and Criminal Justice* and *Fundamentals of Research in Criminology and Criminal Justice* (with Ronet Bachman), *The Practice of Research in Social Work* and *Fundamentals of Social Work Research* (with Ray Engel), and *Research Methods in Education* (with Joseph Check)—he is the author of *Homelessness, Housing, and Mental Illness* and *Organization in a Changing Environment,* coeditor of *Social Neuroscience: Brain, Mind, and Society* and of *The Organizational Response to Social Problems,* and coauthor of *Responding to the Homeless: Policy and Practice.* He has authored and coauthored more than 50 peer-reviewed journal articles as well as many book chapters and research reports on homelessness, service preferences and satisfaction, mental health, organizations, law, and teaching research methods. His research has included a mixed-methods investigation of a public health coordinated care program, a study of community health workers and recruitment for cancer clinical trials, a mixed-methods study of a youth violence reduction program, a randomized trial of a peer support program for homeless dually diagnosed veterans, and a randomized evaluation of housing alternatives for homeless persons diagnosed with severe mental illness, with extramural funding from the National Cancer Institute, the Veterans Health Administration, the National Institute of Mental Health, the Fetzer Institute, and state agencies. His current scholarly foci are the impact of social relations and the social environment on cognitive and community functioning, the meaning of housing and service preferences, and the value of alternative organizational and occupational structures for service delivery. His prior research has included investigation of social factors in legal decisions and admission practices and of influences on job and service satisfaction. Details are available at http://rschutt.wikispaces.umb.edu.

Preface

Are you married, thinking about marrying, or expecting to get married sometime in the future? Whether you answer yes or no, you have a lot of company: Half of all adults in the United States are married and half are not. But since you're in college, you might wonder how your marital expectations compare to those of your educational peers. If so, you'll be interested to learn that 65% of adults aged 25 and older with a 4-year college degree are married, but the marriage rate is only 50% among those with no education beyond high school. And if you have heard the talk about marriage rates declining and look into that, the picture becomes even more interesting: Since 1960, when 72% of U.S. adults were married, the marriage rate has declined much more for those who are less educated than for the college educated (see Exhibit P.1). And do you wonder why? Part of the decline in the overall proportion of married adults is due to marrying later in life, but there is also a continuing increase in the share of Americans who have never married and who live with a partner without marrying. See Parker and Stepler (2017) for an overview of these trends.

Exhibit P.1 The Education Gap in Marriage Continues to Grow



% of U.S. adults ages 25 and older who are married, by education

Note: "Some college" includes those with an associate degree and those who attended college but did not obtain a degree. Adults who are separated are not classified as married.

Source: "As U.S. Marriage Rate Hovers at 50%, Education Gap in Marital Status Widens." Pew Research Center, Washington, DC. (September, 2017). http://www.pewresearch.org/fact-tank/2017/09/14/as-u-s-marriage-rate-hovers-at-50-education-gap-in-marital-status-widens/.

Do you wonder how these findings were obtained? Are you sure the information is trustworthy; that the findings are correct? If you read the summary of the research in the United States by Kim Parker and Renee Stepler (2017), you would learn that these findings come primarily from the American Trends Panel, a survey of 4,971 adults conducted by the Pew Research Center. But is that all you need to know to assess the trustworthiness of these findings? Do you want to know what questions were asked in the survey and how the adults were selected?

You can find answers to these questions in a report on the Pew survey, *The American Trends Panel Survey Methodology* (Pew Research Center 2017a). But these answers will in turn raise more questions: What are "weighted" data? What is a "representative sample"? What is a "margin of sampling error"? How good are the questions asked in the Pew survey (which you can find in another Pew report; Pew Research Center 2017b)? And would you like to know how the marriage rate in the United States compares to those in other countries? You can find those details in the Organisation for Economic Co-operation and Development (OECD) Family Database (OECD 2017) or in *The Sustainable Demographic Dividend,* a report by the Social Trends Institute (2017).

Of course, I have presented only a bit of the evidence about marriage rates and just a few of the conclusions about variation in those rates between social groups and over time; there are many related issues to consider and important questions to ask about the evidence. My intent now is not to give you the whole picture about marriage rates, but to introduce the study of research methods by illustrating how research helps us understand pressing social questions and why we need to learn more about research methods to evaluate the results of such research.

Neither our own limited perceptions nor a few facts gleaned from even reputable sources provide us with a trustworthy basis for understanding the social world. We need systematic methods for investigating our social world that enable us to chart our course through the passions of the moment and to see beyond our own personal experience.

Teaching and Learning Goals

If you see the importance of pursuing answers to questions about the social world the way that the Pew Research Center does, you can understand the importance of investigating the social world. One purpose of this book is to introduce you to social science research methods such as those involved in the Pew study of the education gap in marriage rates and to show how they improve everyday methods of answering our questions about the social world. Each chapter integrates instruction in research methods with investigation of interesting aspects of the social world, such as the use of social networking; the police response to domestic violence; and influences on crime, homelessness, work organizations, health, patterns of democratization, and the response to disasters.

Another purpose of this book is to give you the critical skills necessary to evaluate research. Just "doing research" is not enough. Just reading that some conclusions are "based on a research study" is not sufficient. You must learn to ask many questions before concluding that research-based conclusions are appropriate. What did the researchers set out to investigate? How were people selected for study? What information was collected, and how was it analyzed? Throughout this book, you will learn what questions to ask when critiquing a research study and how to evaluate the answers. You can begin to sharpen your critical teeth on the illustrative studies throughout the book.

A third goal of this book is to train you to actually do research. Substantive examples will help you see how methods are used in practice. Exercises at the end of each chapter give you ways to try different methods alone or in a group. A checklist for research proposals will chart a course when you plan more ambitious studies. But research methods cannot be learned by rote and applied mechanically. Thus, you will learn the benefits and liabilities of each major approach to research and why employing a combination of them is often preferable. You will come to appreciate why the results of particular research studies must always be interpreted within the context of prior research and through the lens of social theory.

Organization of the Book

The way the book is organized reflects my beliefs in making research methods interesting, teaching students how to critique research, and viewing specific research techniques as parts of an integrated research strategy. The text is divided into four sections. The three chapters in the first section, Foundations for Social Research, introduce the why and how of research in general. Chapter 1 shows how research has helped us understand the impact of social networking and changes in social ties. It also introduces some alternative approaches to social research, with a particular emphasis on the contrast between quantitative and qualitative research approaches. Chapter 2 illustrates the basic stages of research with a series of experiments on the police response to domestic violence, it emphasizes the role of theory in guiding research, and it describes the major strategies for research projects. Chapter 3 highlights issues of research ethics by taking you inside Stanley Milgram's research on obedience to authority and by introducing the institutional review boards (IRBs) that examine the ethics of proposed research. The chapter ends by discussing the organization of research proposals.

The next three chapters, Fundamentals of Social Research, discuss how to evaluate the way researchers design their measures (Chapter 4), draw their samples (Chapter 5), and justify their statements about causal connections (Chapter 6). As you learn about these procedures, you will also read about research on substance abuse and gangs, homelessness, and the causes of violence.

In the next section, Basic Social Research Designs, Chapters 7, 8, and 9 present the primary strategies used in quantitative research: collecting data through experiments (Chapter 7) and surveys (Chapter 8) and analyzing data with statistics (Chapter 9). The fascinating research examples in these chapters come from investigations of the causes of interpersonal confrontations, the effects of education on health, and the factors associated with voting. Chapters 10 and 11 then introduce the primary strategies used in collecting qualitative data (including participant observation, intensive interviews, and focus groups), and analyzing the results. You will learn in these two chapters about the response to disasters and the course of social interaction.

The Complex Social Research Designs in Chapters 12 through 16 each can involve multiple methodologies. Chapter 12 introduces the mixed-methods approaches that combine quantitative and qualitative research techniques to improve understanding of the complexities of the social world. Evaluation research, the focus of Chapter 13, can employ experiments, surveys, and qualitative methods to learn about the need for and the effects of social and other types of programs. This chapter begins with an overview of evaluation research on drug abuse prevention programs. Chapter 14 introduces the techniques and challenges of secondary data analysis and the use of Big Data. The online availability of thousands of data sets from social science studies has helped make secondary data analysis —the use of previously collected data to investigate new research questions—the method of choice in many investigations. What is termed *Big Data* is generated when you use social media, find your location with GPS, search the web, purchase goods online, or in other ways create electronic records that can be analyzed by researchers to identify social patterns.

Chapter 15 focuses on historical and comparative methodologies, which can use data obtained with one or more of the primary methods to study processes at regional and societal levels over time and between units; research examples focus on the process of democratization and the bases of social revolutions. This chapter also introduces content analysis, which can be used to good effect in historical and comparative research, as well as in research on issues such as gender roles, to reveal how text or pictures reflect social processes. Chapter 16 finishes up with an overview of the process of and techniques for reporting research results—with special attention to visual representations of data patterns and an overview of the problem of plagiarism and some guidelines for writing about research.

Distinctive Features of the Ninth Edition

The ninth edition of *Investigating the Social World* retains the strengths of previous editions while breaking new ground with the latest developments in research methods, enhanced tools for learning in the text and online, and contemporary, fascinating research findings. Some chapters have been reorganized and renumbered to improve the flow of the text and to better connect related techniques, and there are new examples throughout. You will find many other innovations in approach, coverage, and organization in the ninth edition:

New material reflecting the latest advances in research methods.

The continuing expansion of our social world through the web and the forms of computer-mediated communication it supports continues to stimulate advances in research methods. A new section on systematic reviews in Chapter 2 introduces the availability of the worldwide effort to provide integrated reviews of research on interventions in health, social welfare, education, and other areas. The latest approaches in online surveys and online qualitative research methods have been added to Chapters 8, 10, and 11, while Chapter 14's overview of Big Data methods has been expanded considerably, with attention to geodata and new data sources. One of the key advantages of *Investigating the Social World*—balanced and informed treatment of both qualitative and quantitative research approaches—is thus now complemented with consistent attention to using research methods online as well as offline, so that students can tailor the research methods they use to the research questions they ask and the research opportunities they have. Expected changes in federal standards for human subject research are included in Chapter 3's overview of IRB requirements. New perspectives on data visualization and a new section on research posters have been added to Chapter 16.

Reorganization to improve flow within and between chapters.

The fourth section, complex research designs, has been reorganized to draw attention to the increasing focus on and value of mixed methods, while the overview of meta-analysis has been moved from the last chapter to a final section in Chapter 9, where it complements other topics in quantitative data analysis. Chapter 1 has been streamlined by a reduced focus on alternative research philosophies (with more discussion about them in Chapter 10) with more attention given to the role of values in research. Many chapters have fewer examples in order to improve the flow of the text.

Updated coverage of each research method.

Research methods continue to develop, and new challenges must be overcome as our social world continues to change. Coverage of topics has been updated to reflect increased attention to systematic literature reviews and to provide more guidance for effective literature reviews and the use of new online tools, distinguishing conceptual frameworks and social theories, working as part of a research team, changes in the Federal Policy for the Protection of Human Subjects, new concerns in research ethics, and more examples of unethical practice. I have focused more attention on the expanding role of the online social world, including developments in online experiments and surveys, online qualitative methods, geodata and other forms of Big Data, and visual methods. I have added more guidance on improving survey questions, the use of mixed-mode surveys, and issues in survey ethics. My coverage of qualitative methods includes more guidance on case study methods, an introduction to abductive analysis, new material on narrative analysis, a new section on video ethnography, and new examples of ethical dilemmas, as well as new ways of thinking about and using online approaches. The chapter on mixed methods (now Chapter 12) has been updated, and the chapter on evaluation and policy research (now Chapter 13) includes new perspectives on program theory and evidence-based practice. New sources are provided in the chapter on secondary data analysis (Chapter 14) and coverage of the rapidly growing area of Big Data has been expanded with information on the background of and resources for these approaches. My coverage of these issues reflects the latest insights from SAGE's most recent specialized books on particular research methods as well as new lessons from my own research.

Examples of social research as it occurs in real-world settings.

Most of the *Research That Matters, Questions That Count* examples that introduce each chapter have been updated, and many other examples from recent social research projects have been added. Fascinating examples of research on social ties, domestic violence, crime, and other social issues have been updated and extended from the eighth edition. The examples demonstrate that the exigencies and complexities of real life shape the application of research methods. The *Research in the News* examples have been replaced in each chapter with new vignettes, drawing attention to the role of social research in shaping public discourse and policy.

Web-based instructional aids.

The book's study site includes interactive exercises for the *Research That Matters, Questions That Count* research articles—including those that are new to this edition. These articles were published in SAGE journals and correspond to the primary research topic in each chapter of the text. It is important to spend enough time with these exercises to become very comfortable with the basic research concepts presented. The interactive exercises allow you to learn about research on a range of interesting topics as you practice using the language of research, while the articles are posted so that you can explore the details of the reported research.

Careers and research.

Many of the chapter vignettes about the career of a researcher have been updated. They provide good examples of the value of studying hard and mastering these methods!

Updated links and exercises.

Exercises using websites have been updated and those involving IBM SPSS (Statistical Package for the Social Sciences) Statistics* software have been revised for use with the new 2016 General Social Survey (GSS) data set (a version of this data set is available on the book's study site). Interactive exercises have been updated for the new *Research That Matters* articles, and they link to the SAGE journal articles from which those articles were obtained. The HyperRESEARCH program for qualitative analysis is also available online to facilitate qualitative analyses presented in exercises for Chapter 11.

Aids to effective study.

The many effective study aids included in the previous editions have been updated, as needed. Appendix E, which you will find on the book's study site, presents an annotated list of useful websites.

It is a privilege to be able to share with so many students the results of excellent social science investigations of the social world. If *Investigating the Social World* communicates the excitement of social research and the importance of evaluating carefully the methods we use in that research, then I have succeeded in representing fairly what social scientists do. If this book conveys accurately the latest developments in research methods, it demonstrates that social scientists are themselves committed to evaluating and improving their own methods of investigation. I think it is fair to say that we practice what we preach.

Now you're the judge. I hope that you and your instructor enjoy learning how to investigate the social world and perhaps do some investigating along the way. And I hope you find that the knowledge and (dare I say it?) enthusiasm you develop for social research in this course will serve you well throughout your education, in your career, and in your community.

A Note About Using SPSS[*]

[*] IBM SPSS® Statistics.

To carry out the SPSS exercises at the end of each chapter and in Appendix D (on the book's study site), you must already have SPSS on your computer. The exercises use a subset of the 2016 GSS data set (included on the study site). This data set includes variables on topics such as work, family, gender roles, government institutions, race relations, and politics. Appendix D will get you up and running with IBM SPSS, and you can then spend as much time as you like exploring characteristics and attitudes of Americans. If you are able to use the complete version of SPSS (perhaps in a university computer lab), just download the GSS2016 file and save it on your computer, then start SPSS on your PC, open the GSS2016 file, and begin with the first SPSS exercise in Chapter 1. The GSS2016x file provides all the variables you need for exercises in each chapter except Chapter 9, while the GSS2016y file includes just the variables needed for the examples and exercises in Chapter 9. If you are using the SPSS Student version of SPSS (purchased with this text or separately), you must download the GSS2016x_reduced and GSS2016y_reduced files and use them, as requested, for the SPSS exercises. Alternatively, you could complete many of the SPSS exercises in the text using an online analysis program at the University of California, Berkeley, website (http://sda.berkeley.edu/archive.htm) or at the National Opinion Research Center site (www.norc.uchicago.edu/GSS+Website/). See the book's study site for instructions about this easy approach to statistical analysis.

The study site also includes a subset of the 2002 International Social Survey Program data set. In addition, the GSS website listed subsequently contains documentation files for the GSS2016 and the ISSP2002, as well as the complete GSS2016 data set (but this original dataset does not include constructed variables that are used in some of the exercises).

Ancillaries

SAGE edge™

edge.sagepub.com/schutt9e

SAGE edge offers a robust online environment featuring an impressive array of tools and resources for review, study, and further exploration, keeping both instructors and students on the cutting edge of teaching and learning. SAGE edge content is open access and available on demand. Learning and teaching has never been easier!

SAGE edge for students provides a personalized approach to help students accomplish their coursework goals in an easy-to-use learning environment.

- Mobile-friendly eFlashcards strengthen understanding of key terms and concepts.
- Mobile-friendly practice quizzes allow for independent assessment by students of their mastery of course material.
- Chapter summaries with learning objectives reinforce the most important material.
- EXCLUSIVE! Access to full-text SAGE journal articles that have been carefully selected to support and expand on the concepts presented in each chapter.

SAGE edge for instructors supports teaching by making it easy to integrate quality content and create a rich learning environment for students.

- Test banks provide a diverse range of pre-written options as well as the opportunity to edit any question and/or insert personalized questions to effectively assess students' progress and understanding.
- Sample course syllabi for semester and quarter courses provide suggested models for structuring one's course.
- Editable, chapter-specific PowerPoint® slides offer complete flexibility for creating a multimedia presentation for the course.
- EXCLUSIVE! Access to full-text SAGE journal articles that have been carefully selected to support and expand on the concepts presented in each chapter to encourage students to think critically.
- Multimedia content includes original SAGE videos that appeal to students with different learning styles.
- Lecture notes summarize key concepts by chapter to ease preparation for lectures and class discussions.
- A coursepack provides easy LMS integration.

Acknowledgments

My thanks first to Jeff Lasser, sociology publisher for Sage Publishing. Jeff's consistent support has made it possible for this project to flourish, and his collegiality has even made it all rather fun. Editorial assistant Adeline Wilson also contributed her outstanding talents to the success of this edition and to the quality of the Careers and Research highlights. Book production was managed with great expertise and good cheer by Veronica Stapleton Hooper, while the remarkable Amy Marks again proved herself to be one of the publishing industry's most conscientious and effective copy editors. Assistant content development editor Sarah Dillard and acquisition editor Rachael Leblond artfully managed development of book ancillaries, artwork, and video. I am grateful to work with such talented staff at what has become the world's best publisher in social science.

I also am indebted to the first-rate social scientists Jeff Lasser recruited to provide feedback. Their thoughtful suggestions and cogent insights have helped improve every chapter in the ninth edition. They are

Francis O. Adeola, University of New Orleans
William Augustine, SUNY University at Albany
Qingwen Dong, University of the Pacific
Karen Robinson, California State University San Bernardino
Jennifer Samson, Arkansas Tech University
Pam Tontodonato, Kent State University

The quality of *Investigating the Social World* benefits increasingly from the wisdom and creativity of my coauthors on adaptations for other markets and disciplines, as well as from the pleasure of being part of the support group that we provide each other. My profound gratitude to my Sage coauthors: Ronet Bachman (University of Delaware), Dan Chambliss (Hamilton College), Joe Check (University of Massachusetts Boston), Ray Engel (University of Pittsburgh), and Paul Nestor (University of Massachusetts Boston). And I continue to be grateful for advice shared at my biennial meeting with coauthors at the Harvard Faculty Club by Philip Brenner, Charles DiSogra, Karen Hacker, Sunshine Hillygus, Peter Marsden, Catherine Kohler Riessman, and Robert J. Sampson.

Candace Cantrell, one of our wonderful doctoral students, provided indispensable assistance for the ninth edition, checking websites, updating SPSS exercises and the SPSS appendix, helping identify SAGE articles for the introductory chapter vignettes, finding new *Research in the News* stories, and developing new interactive exercises for this edition. My thanks for the quality of her work and the sophistication of her skills.

Reviewers for the eighth edition were

Karl Besel, Indiana University at Kokomo
Gregory Fulkerson, SUNY Oneonta
Lisa Ann Gittner, Tennessee State University
George Guay, Bridgewater State University
Amy Kroska, University of Oklahoma
Derek Lester, Texas A&M University
Roseanne Macias, California State University at Dominguez Hills
Gina Mann-Delbert, Suffolk University
Tajuana D. Massie, South Carolina State University
Peggy Walsh, Keene University
Greg Weaver, Auburn University

Reviewers for the seventh edition were

Robyn Brown, DePaul University
Jennifer Bulanda, Miami University
Jerry Daday, Western Kentucky University
Marvin Dawkins, University of Miami
Patricia Drentea, University of Alabama at Birmingham
Kenneth Fernandez, University of Nevada, Las Vegas
Elizabeth Monk-Turner, Old Dominion University
David Sanders, Angelo State University
Jimmy Kazaara Tindigarukayo, University of the West Indies
Susan Wurtzburg, University of Hawai'i

Reviewers for the sixth edition were

Von Bakanic, College of Charleston
Marvin Dawkins, University of Miami
Carol Erbes, Old Dominion University
Kenneth E. Fernandez, University of Nevada, Las Vegas
Isaac Heacock, Indiana University, Bloomington
Edward Lascher, California State University, Sacramento
Quan Li, University of Central Florida
Steve McDonald, North Carolina State University
Kevin A. Yoder, University of North Texas

Reviewers for the fifth edition were

James David Ballard, California State University, Northridge
Carl Bankston, Tulane University
Diana Bates, The College of New Jersey
Sandy Cook-Fon, University of Nebraska at Kearny

Christopher Donoghue, William Paterson University
Tricia Mein, University of California at Santa Barbara
Jeanne Mekolichick, Radford University
Kevin Mulvey, George Washington University
Jennifer Parker-Talwar, Pennsylvania State University at Lehigh Valley
Nicholas Parsons, Washington State University
Michael J. Scavio, University of California at Irvine
Shaihid M. Shahidullah, Virginia State University
Tabitha Sharp, Texas Woman's University
John Talmage, University of North Florida
Bill Tillinghas, San Jose State University

Fourth edition reviewers were

Marina A. Adler, University of Maryland, Baltimore
Diane C. Bates, Sam Houston State University
Andrew E. Behrendt, University of Pennsylvania
Robert A. Dentler, University of Massachusetts Boston (Chapter 10)
David H. Folz, University of Tennessee
Christine A. Johnson, Oklahoma State University
Carolyn Liebler, University of Washington
Carol D. Miller, University of Wisconsin–La Crosse
Dan Olson, Indiana University South Bend
Brian J. Stults, University of Florida
John R. Warren, University of Washington
Ken Wilson, East Carolina University

Third edition reviewers were

Emmanuel N. Amadi, Mississippi Valley State University
Doug Anderson, University of Southern Maine
Robert B. Arundale, University of Alaska, Fairbanks
Hee-Je Bak, University of Wisconsin–Madison
Marit Berntson, University of Minnesota
Deborah Bhattacharayya, Wittenbert University
Karen Bradley, University of Central Missouri State
Cynthia J. Buckley, The University of Texas at Austin
J. P. Burnham, Cumberland College
Gerald Charbonneau, Madonna University
Hugh G. Clark, Texas Woman's University
Mark E. Comadena, Illinois State University
John Constantelos, Grand Valley State University
Mary T. Corrigan, Binghamton University

John Eck, University of Cincinnati
Kristin Espinosa, University of Wisconsin–Milwaukee
Kimberly Faust, Fitchburg State College
Kenneth Fidel, DePaul University
Jane Hood, University of New Mexico
Christine Johnson, Oklahoma State University
Joseph Jones, Taylor University
Sean Keenan, Utah State University
Debra Kelley, Longwood College
Kurt Kent, University of Florida
Jan Leighley, Texas A&M University
Joel Lieberman, University of Nevada, Las Vegas
Randall MacIntosh, California State University, Sacramento
Peter J. May, University of Washington
Michael McQuestion, University of Wisconsin–Madison
Bruce Mork, University of Minnesota
Jennifer R. Myhre, University of California, Davis
Zeynep Özgen, Arizona State University
Norah Peters-Davis, Beaver College
Ronald Ramke, High Point University
Adinah Raskas, University of Missouri
Akos Rona-Tas, University of California, San Diego
Therese Seibert, Keene State College
Mark A. Shibley, Southern Oregon University
Pamela J. Shoemaker, Syracuse University
Herbert L. Smith, University of Pennsylvania
Paul C. Smith, Alverno College
Glenna Spitze, University at Albany, State University of New York
Beverly L. Stiles, Midwestern State University
Carolina Tolbert, Kent State University
Tim Wadsworth, University of Washington
Charles Webb, Freed-Hardeman University
Adam Weinberg, Colgate University

Special thanks to Barbara Costello, University of Rhode Island; Nancy B. Miller, University of Akron; and Gi-Wook Shin, University of California, Los Angeles, for their contributions to the third edition.

Second edition reviewers were

Nasrin Abdolali, Long Island University, C. W. Post
Lynda Ames, State University of New York, Plattsburgh
Matthew Archibald, University of Washington

Karen Baird, Purchase College, State University of New York
Kelly Damphousse, Sam Houston State University
Ray Darville, Stephen F. Austin State University
Jana Everett, University of Colorado, Denver
Virginia S. Fink, University of Colorado, Colorado Springs
Jay Hertzog, Valdosta State University
Lin Huff-Corzine, University of Central Florida
Gary Hytrek, University of California, Los Angeles
Debra S. Kelley, Longwood College
Manfred Kuechler, Hunter College (CUNY)
Thomas Linneman, College of William & Mary
Andrew London, Kent State University
Stephanie Luce, University of Wisconsin–Madison
Ronald J. McAllister, Elizabethtown College
Kelly Moore, Barnard College, Columbia University
Kristen Myers, Northern Illinois University
Michael R. Norris, University of Texas, El Paso
Jeffrey Prager, University of California, Los Angeles
Liesl Riddle, University of Texas, Austin
Janet Ruane, Montclair State University
Josephine A. Ruggiero, Providence College
Mary Ann Schwartz, Northeastern Illinois University
Mildred A. Schwartz, University of Illinois, Chicago (Chapter 11)
Gi-Wook Shin, University of California, Los Angeles
Howard Stine, University of Washington
William J. Swart, The University of Kansas
Guang-zhen Wang, Russell Sage College
Shernaaz M. Webster, University of Nevada, Reno
Karin Wilkins, University of Texas, Austin
Keith Yanner, Central College

First edition reviewers were

Catherine Berheide, Skidmore College
Terry Besser, University of Kentucky
Lisa Callahan, Russell Sage College
Herbert L. Costner, formerly of University of Washington
Jack Dison, Arkansas State University
Sandra K. Gill, Gettysburg College
Gary Goreham, North Dakota State University
Barbara Keating, Mankato State University
Bebe Lavin, Kent State University

Scott Long, Indiana University
Elizabeth Morrissey, Frostburg State University
Chandra Muller, University of Texas
G. Nanjundappa, California State University, Fullerton
Josephine Ruggiero, Providence College
Valerie Schwebach, Rice University
Judith Stull, Temple University
Robbyn Wacker, University of Northern Colorado
Daniel S. Ward, Rice University
Greg Weiss, Roanoke College
DeeAnn Wenk, University of Oklahoma

I am also grateful for Kathy Crittenden's support on the first three editions, for the contributions of Herbert L. Costner and Richard Campbell to the first edition, and to Steve Rutter, whose vision and enthusiasm launched the whole project on a successful journey.

The interactive exercises on the website began with a series of exercises that I developed in a project at the University of Massachusetts Boston. They were expanded for the second edition by Tom Linneman and a team of graduate students he directed at the University of Washington—Mark Edwards, Lorella Palazzo, and Tim Wadsworth—and tested by Gary Hytrek and Gi-Wook Shin at the University of California, Los Angeles. My format changes in the exercises for the third edition were tested by my daughter, Julia Schutt. Diane Bates and Matthew Archibald helped revise material for instructors and Judith Richlin-Klonsky revised some examples in Chapter 9 for the third edition. Kate Russell developed a new set of exercises and made many other contributions for the seventh edition. Candace Cantrell and Whitney Gecker added more exercises based on the *Research That Matters* articles in the eighth and ninth editions that are also on the study site.

Philip Brenner provided helpful feedback on Chapter 8 for the ninth edition, as did Reef Youngreen and Phil Kretsedemas for the seventh edition on, respectively, Chapters 3 and 4. Several former faculty, staff, and graduate students at the University of Massachusetts Boston made important contributions to earlier editions: Heather Albertson, Ra'eda Al-Zubi, Bob Dentler and students in his 1993–1994 graduate research methods class, as well as Anne Foxx, Whitney Gecker, Chris Gillespie, Tracey Newman, Megan Reynolds, Kathryn Stoeckert, Tatiana Williams-Rodriguez, and Jeffrey Xavier. Heather Johnson at Northeastern University also contributed to an earlier edition. I continue to be indebted to the many students I have had the opportunity to teach and mentor, at both the undergraduate and graduate levels. In many respects, this book could not have been so successful without the ongoing teaching experiences we have shared. I also share a profound debt to the many social scientists and service professionals with whom I have collaborated in social science research projects.

No scholarly book project can succeed without good library resources, and for these I continue to incur a profound debt to the Harvard University library staff and their extraordinary collection. I also have benefited from the resources maintained by the excellent librarians at the University of Massachusetts Boston.

Again, most important, I thank my wife, Elizabeth, for her love and support, and our daughter, Julia, for the joy she brings to our lives and her own research contributions in the social world.

—*Russell K. Schutt*

Section I Foundations for Social Research

Chapter 1 Science, Society, and Social Research

- Research That Matters, Questions That Count
- Learning About the Social World
- Avoiding Errors in Reasoning About the Social World
  - Observing
  - Generalizing
  - Reasoning
  - Reevaluating
- Research in the News: Social Media and Political Polarization
- Science and Social Science
  - The Scientific Approach
  - Pseudoscience or Science
  - Motives for Social Research
  - Types of Social Research
    - *Descriptive Research*
    - *Exploratory Research*
    - *Explanatory Research*
    - *Evaluation Research*
- Careers and Research
- Strengths and Limitations of Social Research
- Alternative Research Orientations
  - Quantitative and/or Qualitative Methods
  - Philosophical Perspectives
  - Basic Science or Applied Research
  - The Role of Values
- Conclusions

How do you contact your friends and family members who don't live with you? Let us count the ways: Text? E-mail? Social media like Facebook? Phone? In-person? Do you prefer in-person contact or one of these other methods? Do you feel in-person contact is better when you need someone to confide in or are looking for support in response to a personal crisis? What do your parents or grandparents think about this? What about other older acquaintances who didn't grow up using the Internet and cell phones? Do your beliefs correspond with the conclusions of Roger Patulny and Claire Seaman?

The Internet, cell phones, and all the interrelated forms of communication they support—e-mail, texting, social media, and video chat, just to name a few—have added new forms of social connection across the globe in the past four decades. By the end of March 2017, 49.7% of the total world population of 7,519,028,970 was connected in some way to the

Internet—an increase of 936% since 2000. Across continents, the percentage connected ranged from highs of 88.1% in North America and 77.4% in Europe to 45.2% in Asia and just 28.3% in Africa (Internet World Stats 2017). As you can imagine, many social scientists have wondered how these developments have affected our lives.

Research That Matters, Questions That Count

Are face-to-face contacts between people being displaced by modern indirect ("mediated") contact through texting, Skype, social media, e-mails, or cell phones? And if so, does it matter? Do people feel less supported when they communicate indirectly compared to when their social contacts are physically present? Since the spread of cell phones and the development of the Internet in the 1980s, social scientists have been concerned with the impact of these new forms of mediated connections on the quantity and quality of social interaction. Professor Roger Patulny and PhD candidate Claire Seaman at the University of Wollongong in Australia investigated these questions with data collected in the Australian Bureau of Statistics' (ABS's) General Social Survey. The procedures for the ABS-GSS involve in-person interviews with over 10,000 Australians selected from throughout Australia so that they are very similar to the total population. In the years studied by Patulny and Seaman (2002, 2006, 2010), the GSS included questions about frequency and methods of contacting family or friends (who respondents were not living with). There were also survey questions about feelings of social support, as well as personal characteristics like age and education. The researchers found that face-to-face contact had declined and use of the new mediated forms of social contact had increased, but there had been no general decline in feelings of having social support. However, there were some disadvantages in frequency of contact and feelings of social support among older men and in relation to having less education or less income.

1. Are your impressions of the impact of modern mediated forms of communication on face-to-face contact similar to the conclusions of Patulny and Seaman?
2. Do you think their findings would have differed in other countries?

In this chapter, you will learn more about the methods that Patulny and Seaman used as well as about other studies of social interaction and mediated forms of communication. By the end of the chapter, you will have a good overview of the approach that researchers use to study social issues like these and others. You can learn more about these approaches by reading the 2017 *Journal of Sociology* article by Roger Patulny and Claire Seaman at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 1 at edge.sagepub.com/schutt9e.

Patulny, Roger and Claire Seaman. 2017. "'I'll Just Text You': Is Face-to-Face Social Contact Declining in a Mediated World?" *Journal of Sociology* 53(2):285–302.

That's where social researchers begin, with questions about the social world and a desire to answer them. What makes social research different from the ordinary process of thinking about our experiences is a focus on broader questions that involve people outside our immediate experience, questions about why things happen that we may not otherwise consider, and the use of systematic research methods to answer those questions. The research by Patulny and Seaman (2017) illustrates this approach. Their conclusions about Australia were similar to those of Keith N. Hampton, Lauren Sessions Goulet, Lee Rainie, and Kirsten Purcell (2011) about the United States. Hampton and colleagues found that Internet usage by adults complemented their other social ties, rather than displacing them, based on responses to the Social Networking Sites and Facebook Survey (funded by the Pew Research Center's Internet & American Life Project).

In this chapter, we focus on questions about Internet use, social networking services, and social ties. As we do so, I hope to convince you that the use of research methods to investigate questions about the social world results in knowledge that can be more important, more trustworthy, and more useful than can personal opinions or individual experiences. You will learn how social scientists' investigations are helpful in answering questions about social ties and the impact of the Internet on these ties. You will also learn about the challenges that researchers confront. By the chapter's end, you should know what is "scientific" in social science and appreciate how the methods of science can help us understand the problems of society.

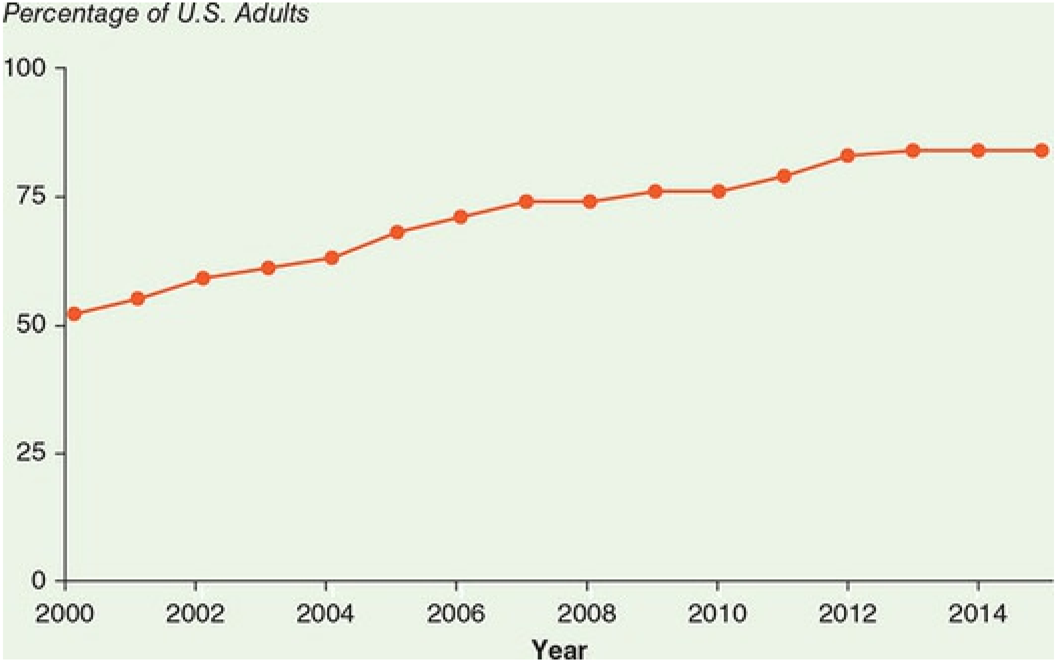
Learning About the Social World

We can get a sense of how sociologists and other social scientists investigate the social world by reviewing some questions that social researchers have asked about the Internet and social ties and the ways they have answered those questions.

1. What percentage of Americans are connected to the Internet?

   The Pew Research Center's surveys have found that Internet use in the United States has risen rapidly from 52% of U.S. adults in 2000 to 84% in 2015 (Perrin and Dugan 2015) (see Exhibit 1.1).
2. How does Internet use vary across social groups?

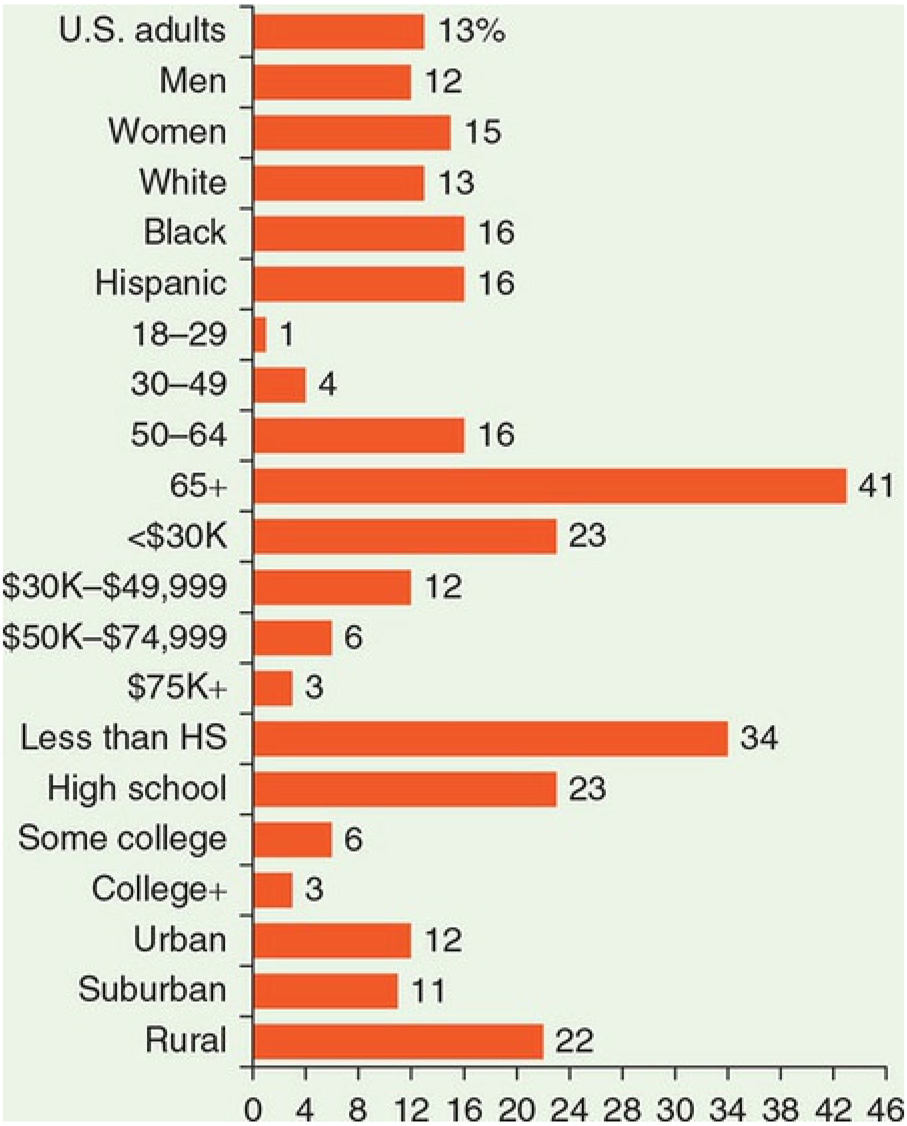
Exhibit 1.1 Percentage of U.S. Adults Who Use the Internet, 2000–2015



Source: "Americans' Internet Access: 2000-2015." Pew Research Center, Washington, DC. (June, 2015.) http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/.

Pew's surveys have also revealed differences in Internet use between social groups. The percentage of U.S. adults who were not online in 2016 (13% overall) was similar between men and women, and by race, but varied dramatically by age—from a low of 1% of those aged 18–29 to a high of 41% among those 65 or older—and by income, education, and location (Anderson and Perrin 2016) (see Exhibit 1.2).

Exhibit 1.2 Percentage of Individuals Not Using Internet, by Personal Characteristics



Source: "13% of Americans Don't Use the Internet. Who Are They?" Pew Research Center, Washington, DC. (September, 2016.) http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/.

3. Does Internet use interfere with the maintenance of social ties?

It doesn't seem so. You have already seen evidence for this conclusion for Australia (Patulny and Seaman 2017). Also in the United States, the extent of social isolation —people not having anyone to confide in—did not change much from 1985 (8%) to 2008 (12%), although there was an apparent uptick in 2004 due to changes in the survey design (Fischer 2009; Hampton et al. 2009; Marsden 1987; McPherson, Smith-Lovin, and Brashears 2006:358; Paika and Sanchagrina 2013) (see Exhibit 1.3). Individuals who use the Internet tend to have larger and more diverse social networks than others do and are about as likely as those who do not use the Internet to participate in community activities.

4. Does wireless access (Wi-Fi) in such public places as Starbucks decrease social interaction among customers?

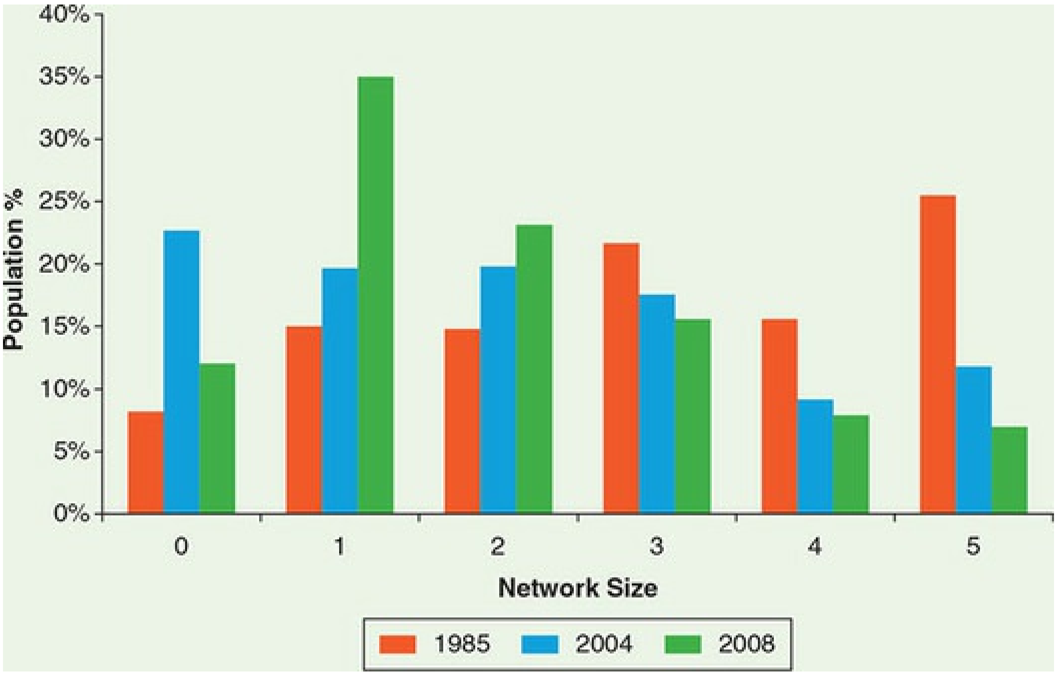
   Hampton and Gupta (2008) observed Internet use in coffee shops with wireless access in two cities and concluded that there were two types of Wi-Fi users: some who used their Internet connection to create a secondary work office and others who used their Internet connection as a tool for meeting others in the coffee shop. What this means is that Wi-Fi was associated with less social interaction among some customers, but more interaction among others.

5. Do cell phones and e-mail tend to hinder the development of strong social ties?

   Based on surveys in Norway and Denmark, Rich Ling and Gitte Stald (2010) concluded that mobile phones increase social ties among close friends and family members, whereas e-mail communication tends to decrease the intensity of our focus on close friends and family members. Other research by the Pew Research Center has identified positive effects of the Internet and e-mail on social ties (Boase et al. 2006).

   Did your personal experiences lead you to expect different answers to these questions? You have just learned that younger people use the Internet more than do older people. Does this variability lead you to be cautious about using your own experience as a basis for estimating the behavior of others (#2)? Have you heard others complain about the effect of the Internet on the maintenance of social ties? Is it safe to draw general conclusions from this anecdotal evidence (#3)? Have you been sensitive to the effects of surroundings and of mode of communication on different people (#4 and #5)?

Exhibit 1.3 Change in Social Network Size



Source: Hampton, Keith N., Lauren Sessions Goulet, Eun Ja Her, and Lee Rainie. November 2009. “Social Isolation and New Technology: How the Internet and Mobile Phones Impact Americans’ Social Networks.” Washington, DC: Pew Internet & American Life Project.

We cannot avoid asking questions about our complex social world or trying to make sense of our position in it. Actually, the more that you begin to “think like a social scientist,” the more such questions will come to mind—and that’s a good thing! But as you’ve just seen, in our everyday reasoning about the social world, our own prior experiences and orientations can have a major influence on what we perceive and how we interpret these perceptions. As a result, one person may see someone posting a message on Facebook as being typical of what’s wrong with modern society, but another person may see the same individual as helping people “get connected” with others. We need to move beyond first impressions and gut reactions to more systematic methods of investigation.

Avoiding Errors in Reasoning About the Social World

How can we avoid errors rooted in the particularities of our own backgrounds and improve our reasoning about the social world? First, let's identify the different processes involved in learning about the social world and the types of errors that can result as we reason about the social world.

When we learn about the social world, we engage in one or more of four processes: (1) *observing* through our five senses (seeing, hearing, feeling, tasting, or smelling); (2) *generalizing* from what we have observed to other times, places, or people; (3) *reasoning* about the connections between different things that we have observed; and (4) *reevaluating* our understanding of the social world on the basis of these processes. It is easy to make mistakes with each of these processes.

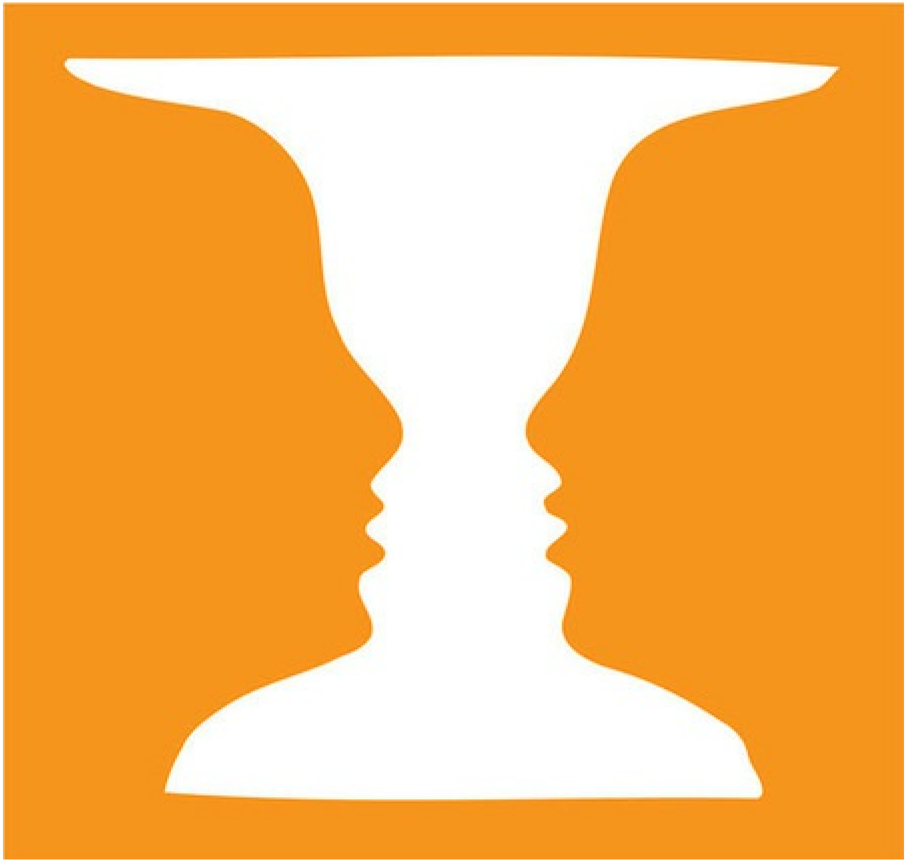
My favorite example of the errors in reasoning that occur in the nonscientific, unreflective discourse about the social world that we hear on a daily basis comes from a letter to famous advice columnist Ann Landers. The letter was written by someone who had just moved with her two cats from the city to a house in the country. In the city, she had not let her cats outside and felt guilty about confining them. When they arrived in the country, she threw her back door open. Her two cats cautiously went to the door and looked outside for a while, then returned to the living room and lay down. Her conclusion was that people shouldn't feel guilty about keeping their cats indoors—even when they have the chance, cats don't really want to play outside.

Do you see this person's errors in her approach to

- *Observing?* She observed the cats at the outside door only once.
- *Generalizing?* She observed only two cats, both of which previously were confined indoors.
- *Reasoning?* She assumed that others feel guilty about keeping their cats indoors and that cats are motivated by feelings about opportunities to play.
- *Reevaluating?* She was quick to conclude that she had no need to change her approach to the cats.

You don't have to be a scientist or use sophisticated research techniques to avoid these four errors in reasoning. If you recognize these errors for what they are and make a conscious effort to avoid them, you can improve your own reasoning about the social world. In the process, you will also be implementing the admonishments of your parents (or minister, teacher, or any other adviser) to avoid stereotyping people, to avoid jumping to conclusions, and to look at the big picture. These are the same errors that the methods of social science are designed to help us avoid.

Exhibit 1.4 An Optical Illusion



Observing

One common mistake in learning about the social world is selective observation—choosing to look only at things that are in line with our preferences or beliefs. When we are inclined to criticize individuals or institutions, it is all too easy to notice their every failure. For example, if we are convinced in advance that all heavy Internet users are antisocial, we can find many confirming instances. But what about elderly people who serve as Internet pen pals for grade-school children? Doctors who exchange views on medical developments? Therapists who deliver online counseling? Couples who maintain their relationship when working in faraway cities? If we acknowledge only the instances that confirm our predispositions, we are victims of our own selective observation.

Selective observation: Choosing to look only at things that are in line with our preferences or beliefs.

Our observations can also simply be inaccurate. If, after a quick glance around the computer lab, you think there are 14 students present, when there are actually 17, you have made an inaccurate observation. If you hear a speaker say that "for the oppressed, the flogging never really stops," when what she said was, "For the obsessed, the blogging never really stops" (Hafner 2004), you have made an inaccurate observation.

Inaccurate observation: An observation based on faulty perceptions of empirical reality.

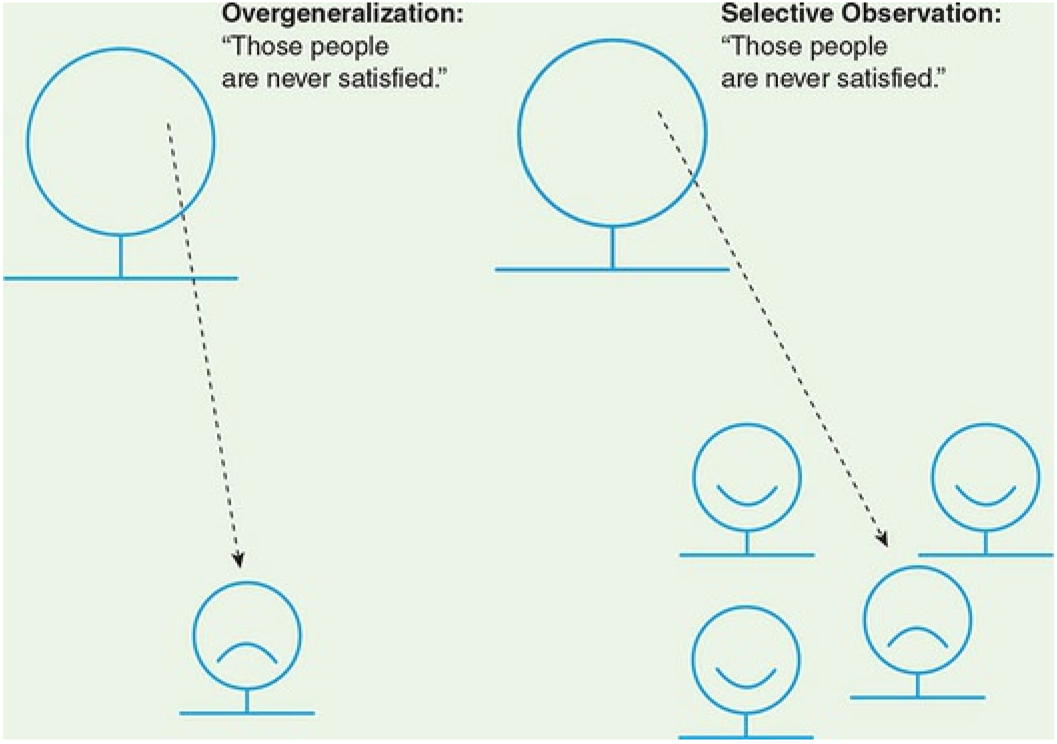
Such errors occur often in casual conversation and in everyday observation of the world around us. In fact, our perceptions do not provide a direct window onto the world around us, for what we think we have sensed is not necessarily what we have seen (or heard, smelled, felt, or tasted). Even when our senses are functioning fully, our minds have to interpret what we have sensed (Humphrey 1992). The optical illusion in Exhibit 1.4, which can be viewed as either two faces or a vase, should help you realize that perceptions involve interpretations. Different observers may perceive the same situation differently because they interpret it differently.

Generalizing

Overgeneralization occurs when we conclude that what we have observed or what we know to be true for some cases is true for all or most cases (see Exhibit 1.5). We are always drawing conclusions about people and social processes from our own interactions with them and perceptions of them, but sometimes we forget that our experiences are limited. The social (and natural) world is, after all, a complex place. We have the ability (and inclination) to interact with just a small fraction of the individuals who inhabit the social world, especially within a limited span of time. Thanks to the Internet, social media, and the practice of "blogging" (i.e., posting personal ruminations on websites), we can easily find many examples of overgeneralization in people's thoughts about the social world. Here's one posted by a frequent blogger who was called for jury duty (http://busblog.tonypierce.com/2005/06/yesterday-i-had-to-go-to-jury-duty-to.html, posted on June 17, 2005):

Overgeneralization: When we unjustifiably conclude that what is true for some cases is true for all cases.

Exhibit 1.5 The Difference Between Selective Observation and Overgeneralization



> yesterday i had to go to jury duty to perform my civil duty. *unlike most people* i enjoy jury duty because i find the whole legal process fascinating, especially when

> its unfolding right in front of you and you get to help decide yay or nay.

Do you know what the majority of people think about jury duty? According to a Harris Poll, 75% of Americans consider jury service to be a privilege (Grey 2005), so the blogger's generalization about "most people" is not correct. Do you ever find yourself making a quick overgeneralization like that?

Reasoning

When we prematurely jump to conclusions or argue on the basis of invalid assumptions, we are using illogical reasoning. An Internet blogger posted a conclusion about the cause of the tsunami wave that devastated part of Indonesia in 2004 (cited in Schwartz 2005):

> Since we know that the atmosphere has become contaminated by all the atomic testing, space stuff, electronic stuff, earth pollutants, etc., is it logical to wonder if: Perhaps the "bones" of our earth where this earthquake spawned have also been affected?

Is that logical? Another blogger soon responded with an explanation of plate tectonics: "The floor of the Indian Ocean slid over part of the Pacific Ocean" (Schwartz 2005:A9). The earth's crust moves no matter what people do!

It is not always so easy to spot illogical reasoning. For example, by September 2016 only 13% of American households reported not using the Internet (File 2013). Would it be reasonable to propose that they don't participate in the "information revolution" simply because they think it is a waste of time? In fact, many low-income households lack the financial resources to buy a computer or maintain an online account and so they use the Internet much less frequently—as you can see in Exhibit 1.2 (Rainie and Horrigan 2005:63). Conversely, an unquestioned assumption that everyone wants to connect to the Internet may overlook some important considerations; for example, 17% of nonusers of the Internet said in 2002 that the Internet has made the world a worse place, so they may not use it because they don't like what they believe to be its effects (UCLA Center for Communication Policy 2003:78). Logic that seems impeccable to one person can seem twisted to another.

Reevaluating

Resistance to change, the reluctance to reevaluate our ideas in light of new information, may occur for several reasons:

Illogical reasoning: When we prematurely jump to conclusions or argue on the basis of invalid assumptions.

Resistance to change: The reluctance to change our ideas in light of new information.

- *Ego-based commitments.* We all learn to greet with some skepticism the claims by leaders of companies, schools, agencies, and so on that people in their organization are happy, that revenues are growing, and that services are being delivered in the best possible way. We know how tempting it is to make statements about the social world that conform to our own needs rather than to the observable facts. It can also be difficult to admit that we were wrong once we have staked out a position on an issue. Barry Wellman (Boase et al. 2006:1) recounts a call from a reporter after the death of four "cyber addicts." The reporter was already committed to the explanation that computer use had caused the four deaths; now, he just wanted an appropriate quote from a computer-use expert, such as Wellman. But the interview didn't last long:

  > The reporter lost interest when Wellman pointed out that other causes might be involved, that "addicts" were a low percentage of users, and that no one worries about "neighboring addicts" who chat daily in their front yards. (Boase et al. 2006:1)

- *Excessive devotion to tradition.* Some degree of devotion to tradition is necessary for the predictable functioning of society. Social life can be richer and more meaningful if it is allowed to flow along the paths charted by those who have preceded us. Some skepticism about the potential for online learning once served as a healthy antidote to unrealistic expectations of widespread student enthusiasm (Bray 1999). But too much devotion to tradition can stifle adaptation to changing circumstances. When we distort our observations or alter our reasoning so that we can maintain beliefs that "were good enough for my grandfather, so they're good enough for me," we hinder our ability to accept new findings and develop new knowledge. Of course, there was nothing "traditional" about maintaining social ties through e-mail when this first became possible in the late 20th century. Many social commentators assumed that the result of increasing communication by e-mail would be fewer social ties maintained through phone calls and personal contact. As a result, it was claimed, the social world would be impoverished. But subsequent research indicated that people who used e-mail more also kept in touch with others more in person and by phone (Benkler 2006:356; Boase et al. 2006).

- *Uncritical agreement with authority.* If we do not have the courage to evaluate critically the ideas of those in positions of authority, we will have little basis for complaint if they exercise their authority over us in ways we don't like. And, if we do not allow new discoveries to challenge our beliefs, our understanding of the social world will remain limited. Was it partly uncritical agreement with computer industry authorities that led so many to utopian visions for the future of the Internet? "Entrepreneurs saw it as a way to get rich, policy makers thought it could remake society, and business people hoped that online sales would make stock prices soar. Pundits preached the gospel of the new Internet millennium" (Wellman 2004:25).

Now take just a minute to reexamine the issues about social ties and Internet use that I discussed earlier. Did you grasp at a simple explanation even though reality is far more complex? Did your own ego and feelings about your similarities to or differences from others influence your beliefs? Did you weigh carefully the opinions of authorities who decry the decline of "community"? Could knowledge of research methods help improve your own understanding of the social world? Do you see some of the challenges social science faces?

Science and Social Science

The scientific approach is designed to reduce greatly these potential sources of error in everyday reasoning. Science relies on logical and systematic methods to answer questions, and it does so in a way that allows others to inspect and evaluate its methods. In this way, scientific research develops a body of knowledge that is continually refined, as beliefs are rejected or confirmed on the basis of testing empirical evidence.

Science: A set of logical, systematic, documented methods for investigating nature and natural processes; the knowledge produced by these investigations.

The Scientific Approach

The sciences—physical (e.g., physics), natural (e.g., biology), social (e.g., sociology)—share an approach that developed over centuries and was formalized in the 17th and 18th centuries by philosophers such as Francis Bacon (1561–1626) and early scientists, including Galileo (1564–1642) and Newton (1643–1727). The central element of this approach (often called the "scientific method" or the "scientific attitude") is investigating phenomena in the world by testing ideas about them with observations—empirical data—of those phenomena. Will two stones of different weight fall at the same or different speeds in a vacuum? Test it out—by dropping them (Galileo). Is white light composed of a mixture of different colors? Test it out—with a prism (Newton).

In the News Research in the News: Social Media and Political Polarization

For Further Thought?

Is the growing importance of social media responsible for increasing political polarization in the United States? After all, social media helps people restrict their information to news with the slant they prefer and their social connections to like-minded partisans. But using data from the American National Election Studies, economics professors at Brown and Stanford Universities found that polarization has been most extreme among older Americans—the age group that is least likely to use social media. So there seems to be more behind polarization than just the use of social media.

1. What else do you think might explain increasing political polarization?
2. In addition to surveys, what data sources could you use to study political polarization?

News source: Bromwich, Jonah Engel. 2017. "Social Media Is Not Contributing Significantly to Political Polarization, Paper Says." *The New York Times,* April 13.

Many scientists accept other elements as important in the scientific approach (Grinnell 1992; Popper 1968; Wallace 1983):

- *Falsifiability.* Scientific ideas can be put to a test and potentially shown to be false.
- *Theoretical.* Science seeks general explanations for phenomena.
- *Empiricism.* Science focuses on phenomena in the real world that can be observed directly or indirectly.
- *Objectivity.* The goal of science is objective assessment of evidence—freedom from bias due to personal background or pressures—but what this means in practice is a commitment to "intersubjectivity"—accepting as scientifically trustworthy only evidence that receives support from other scientists.
- *Community.* Scientific research is conducted in a community of scientists who share and challenge each other's findings and beliefs.
- *Simplicity.* A proposed explanation is preferred over others if—other things being equal—it is simpler.

We will return to the scientific approach in more detail in the next chapter.

Social science relies on scientific methods to investigate individuals, societies, and social processes. It is important to realize that when we apply scientific methods to understanding ourselves, we often engage in activities—asking questions, observing social groups, or counting people—that are similar to things we do in our everyday lives. However, social scientists develop, refine, apply, and report their understanding of the social world more systematically, or "scientifically," than Joanna Q. Public does:

- Social science research methods can reduce the likelihood of overgeneralization by using systematic procedures for selecting individuals or groups to study that are representative of the individuals or groups to which we want to generalize.
- To avoid illogical reasoning, social researchers use explicit criteria for identifying causes and for determining whether these criteria are met in a particular instance.
- Social science methods can reduce the risk of selective or inaccurate observation by requiring that we measure and sample phenomena systematically.
- Because they require that we base our beliefs on evidence that can be examined and critiqued by others, scientific methods lessen the tendency to develop answers about the social world from ego-based commitments, excessive devotion to tradition, or unquestioning respect for authority.

Even as you learn to appreciate the value of social science methods, however, you shouldn't forget that *social* scientists face three specific challenges:

1. The objects of our research are people like us, so biases rooted in our personal experiences and relationships are more likely to influence our conclusions.
2. Those we study can evaluate us, even as we study them. As a result, subjects' decisions to "tell us what they think we want to hear" or, alternatively, to refuse to cooperate in our investigations can produce misleading evidence.
3. In physics or chemistry, research subjects (objects and substances) may be treated to extreme conditions and then discarded when they are no longer useful. However, social (and medical) scientists must concern themselves with the way their human subjects are treated in the course of research (much could also be said about research on animals, but this isn't the place for that).

We must never be so impressed with the use of scientific methods in investigations of the social world that we forget to evaluate carefully the quality of the resulting evidence. And we cannot ignore the need always to treat people ethically, even when that involves restrictions on the manipulations in our experiments, the questions in our surveys, or the observations in our field studies.

Social science: The use of scientific methods to investigate individuals, societies, and social processes; the knowledge produced by these investigations.

Pseudoscience or Science

We must also be on guard against our natural tendency to be impressed with knowledge that is justified with what sounds like scientific evidence, but which has not really been tested. Pseudoscience claims are not always easy to identify, and many people believe them (Shermer 1997:33).

> Pseudoscience: Claims presented so that they appear scientific even though they lack supporting evidence and plausibility.

Are you surprised that more than half of Americans believe in astrology, with all its charts and numbers and references to stars and planets, even though astrological predictions have been tested and found baseless (Shermer 1997:26)? Are any of your beliefs based on pseudoscience?

Motives for Social Research

Similar to you, social scientists have friends and family, observe other persons' social ties, and try to make sense of what they experience and observe. For most, that's the end of it. But for some social scientists, the quality and impact of social ties has become a major research focus. What motivates selection of this or any other particular research focus? Usually, it's one or more of the following reasons:

- *Policy motivations.* Many government agencies, elected officials, and private organizations seek better descriptions of social ties in the modern world so that they can identify unmet strains in communities, deficits in organizations, or marketing opportunities. Public officials may need information for planning zoning restrictions in residential neighborhoods. Law enforcement agencies may seek to track the connections between criminal gangs and the effect of social cohesion on the crime rate. Military leaders may seek to strengthen unit cohesion. These policy guidance and program management needs can stimulate numerous research projects. As Kathleen Cooper and Nancy Victory (2002) said in their foreword to a U.S. Department of Commerce report on the Census Bureau's survey of Internet use,

  > This information will be useful to a wide variety of policymakers and service providers . . . help all of us determine how we can reach Americans more effectively and take maximum advantage of the opportunities available through new information technologies. (p. iii)

- *Academic motivations.* Questions about changing social relations have stimulated much academic social science. One hundred years ago, Émile Durkheim (1951) linked social processes stemming from urbanization and industrialization to a higher rate of suicide. Fifty years ago, David Reisman (1950/1969) considered whether the growing role of the mass media, among other changes, was leading Americans to become a "lonely crowd." Similar to this earlier research, contemporary investigations of the effect of computers and the Internet are often motivated by a desire to understand influences on the strength and meaning of social bonds. Does a "virtual community" in cyberspace perform the same functions as face-to-face social relationships (Norris 2004)? The desire to understand better how the social world works is motivation enough for many social scientists (Hampton and Wellman 2001):

  > It is time to move from speculation to evidence. . . . The growth of computer-mediated communication (CMC) introduces a new means of

> social contact with the potential to affect many aspects of personal communities. . . . This article examines . . . how this technology affected contact and support. (pp. 477, 479)

- *Personal motivations.* Some social scientists who conduct research on social ties feel that by doing so they can help improve the quality of communities, the effectiveness of organizations, or the physical and mental health of many social groups. Social scientists may become interested in social ties as a result of exposure to problems in the social world, or by watching the challenges their children face in school, or for many other reasons, including finding themselves without many friends after a career move. Exhibit 1.6 displays a photograph of Mexican immigrants living in poverty. Can you imagine a college student, in later years, developing an interest in research on poverty in other countries as a result of a study abroad experience that exposed her to such sights?

Exhibit 1.6 An Impoverished Mexican Immigrant Family

Source: National Geographic Creative/Alamy Stock Photo.

Types of Social Research

Whatever the motives, there are four types of social research projects. This section illustrates each type with projects from the large body of research about various aspects of social ties.

Descriptive Research

Defining and describing social phenomena of interest is a part of almost any research investigation, but descriptive research is often the primary focus of the first research about some issue. Descriptive questions asked in research on social ties have included the following: What is the level of particular types of social ties in America (McPherson et al. 2006)? How has the frequency of different forms of social contact changed over time in Australia (Patulny and Seaman 2017)? What social and cultural patterns characterize disadvantaged neighborhoods (Harding 2007)? Measurement (the topic of Chapter 4) and sampling (Chapter 5) are central concerns in descriptive research. Survey research (Chapter 8) is often used for descriptive purposes. Some comparative research also has a descriptive purpose (Chapter 15).

Descriptive research: Research in which social phenomena are defined and described.

Example: Comings and goings on Facebook?

Lee Rainie, director of the Pew Internet Project, and his colleagues Aaron Smith and Maeve Duggan (2013) sought to describe the frequency with which Americans stopped using Facebook and the reasons they did so. To investigate this issue, they surveyed 1,006 American adults by phone and asked them such questions as,

> Do you ever use Facebook?

and

> Have you ever voluntarily taken a break from using Facebook for a period of several weeks or more?

They found that two thirds of American adults who use the Internet also use Facebook and that most (61%) say they have voluntarily taken a break from using Facebook at some time for at least several weeks (Rainie et al. 2013). Rainie et al. also found that one fifth of Internet users said they had once used Facebook but no longer do so, whereas almost 1 in 10 Internet users who had not used Facebook were interested in doing so. Among those who had stopped using Facebook at some point, reasons for the “break” included not having enough time, lacking interest in the site, not seeing valuable content, and disliking gossiping by their friends.

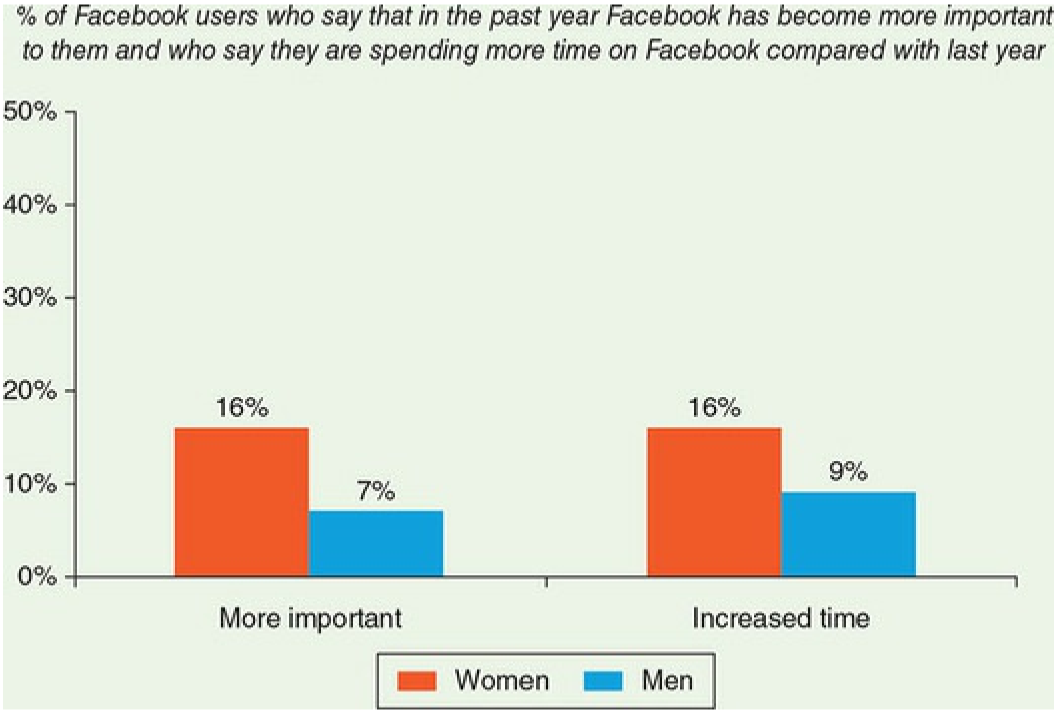
As indicated in Exhibit 1.7, Rainie et al. also found that women were more likely to report increased interest in Facebook and to expect increased use in the next year.

Exploratory Research

Exploratory research seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them. The goal is to learn “What is going on here?” and to investigate social phenomena without explicit expectations. This purpose is associated with the use of methods that capture large amounts of relatively unstructured information or that take a field of inquiry in a new direction. For example, researchers investigating social ties occurring through the Internet have had to reexamine the meaning of “community,” asking whether cyberspace interactions can constitute a community that is seen as “real and essential” to participants (Fox and Roberts 1999:644). “How is identity—true or counterfeit—established in online communities?” asked Peter Kollock and Marc Smith (1999:9). Exploratory research such as this frequently involves qualitative methods, which are the focus of Chapters 10 and 11, as well as special sections in many other chapters.

Exploratory research: Research that seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them.

Exhibit 1.7 The Value of Facebook



Source: Pew Research Center’s Internet & American Life Project Omnibus Survey, conducted December 13 to 16, 2012, on landline and cell phones.

Note: N for male Facebook users = 233. N for female Facebook users = 292.

Example: How do Internet resources help elderly persons manage heart conditions?

The Internet provides a “space where disparate individuals can find mutual solace and exchange information within a common community of interest” (Loader et al. 2002:53). British social scientists Sally Lindsay, Simon Smith, Frances Bell, and Paul Bellaby (2007) were impressed with the potential of Internet-based health resources and wondered how elderly persons who use the Internet to help manage heart conditions would feel about their experiences.

Lindsay and her colleagues decided to explore this question by introducing a small group of older men to computers and the Internet and then letting them discuss their experiences with using the Internet for the next 3 years. Lindsay and her colleagues conducted interviews and a group discussion with their participants. Using a systematic process, the researchers identified different key themes in the transcripts of the interviews and discussions. For example, one man said: “There’s a lot of information there. It makes you feel a lot better. It takes a lot of the fear away.” (Lindsay et al. 2007:103).

Lindsay et al. (2007:107) concluded the Internet provided these new users with both more knowledge and greater social support in dealing with their health problems.

Explanatory Research

Many people consider explanation the premier goal of any science. Explanatory research seeks to identify the causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon. Internet researchers adopted explanation as a goal when they began to ask such questions as "Does the Internet increase, decrease, or supplement social capital?" (Patulny and Seaman 2017; Wellman et al. 2001). "Do students who meet through Internet interaction like each other more than those who meet face-to-face"? (Bargh, McKenna, and Fitzsimons 2002:41). And "how [does] the Internet affect the role and use of the traditional media?" (Nie and Erbring 2002:276). I focus on ways of identifying causal effects in Chapter 6. Explanatory research often involves experiments (see Chapter 7) or surveys (see Chapter 8), both of which are most likely to use quantitative methods.

Example: What effect does Internet use have on social relations?

Jeffrey Boase, John B. Horrigan, Barry Wellman, and Lee Rainie (2006), sociologists at the University of Toronto at the time (Boase and Wellman) and researchers at the Pew Internet Project (Horrigan and Rainie), sought to understand how the Internet is affecting community life in general and the maintenance of social ties in particular. For this purpose, they analyzed data from two phone surveys, conducted in 2004 and 2005, of 4,401 Americans. The surveys included questions about Internet use, social ties, help seeking, and decision making.

Boase and his coauthors (2006) found that the Internet and e-mail help people maintain dispersed social networks and do not conflict with the maintenance of social ties in the local community involving personal or phone contact. The researchers actually found that people who have more in-person and phone connections also tend to use the Internet more. The social value of the Internet is also increased because it is used to seek help and make decisions at important times. How do these conclusions compare with those of Patulny and Seaman (2017)? What might explain the differences?

Explanatory research: Research that seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon.

Evaluation Research

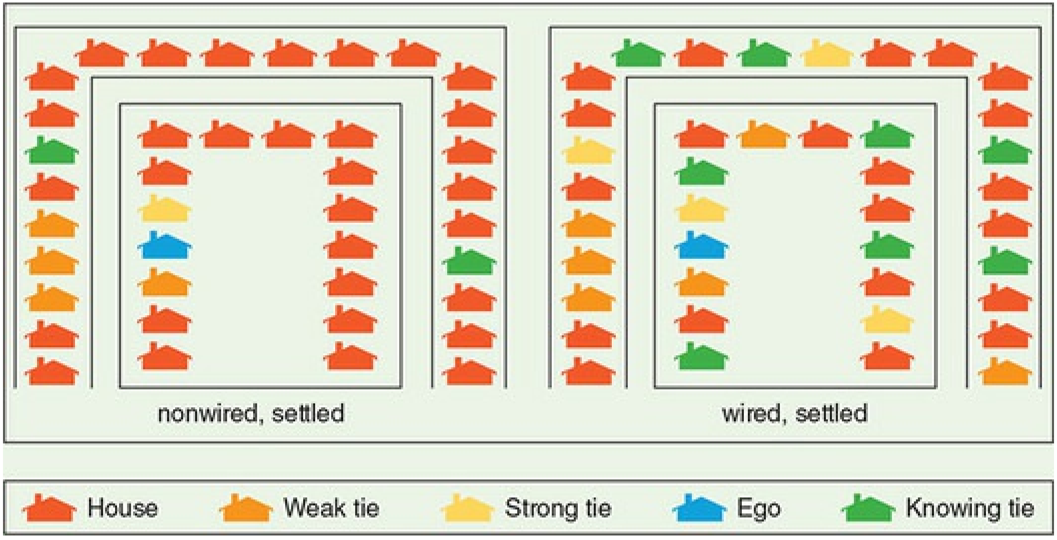
Evaluation research seeks to determine the effects of programs, policies, or other efforts to affect social patterns, whether by government agencies, private nonprofits, or for-profit businesses. This is a type of explanatory research because it deals with cause and effect, but it differs from other forms of explanatory research because evaluation research focuses on one type of cause: programs, policies, and other conscious efforts to create change (Lewis-Beck, Bryman, and Liao 2004:337). This focus raises some issues that are not relevant in other types of explanatory research. Concern regarding the potential impact of alternative policies regarding the Internet provided an impetus for new evaluation research. Chapter 13 introduces evaluation research.

Example: Does high-speed Internet access change community life?

In a new suburban Toronto community termed "Netville," developers connected all homes with a high-speed cable and appropriate devices for Internet access. Sociologists Barry Wellman and Keith Hampton (1999) used this arrangement to evaluate the impact of Internet access on social relations. They surveyed Netville residents who were connected to the Internet and compared them with residents who had not activated their computer connections. Hampton actually lived in Netville for 2 years, participating in community events, taking notes on social interaction and discussion list postings, and conducting surveys (Hampton and Wellman 2000).

Hampton and Wellman (2001) found that Internet access increased residents' social relations ("Ego" in Exhibit 1.8) with other households, resulting in a larger and less geographically concentrated circle of friends. Information about home repair and other personal and community topics and residents' service needs were exchanged over the Internet. Sensitive personal topics, however, were discussed offline. Moreover, the wired residents were not more likely to physically visit other residents than the nonwired residents (Hampton 2003:422).

Exhibit 1.8 The Development of Social Ties in a New Wired and Nonwired Neighborhood



Source: Hampton and Wellman 2000. Reprinted with permission.

Evaluation research: Research that describes or identifies the impact of social policies and programs.

Careers and Research

Jessica LeBlanc, Research Assistant

Jessica LeBlanc majored in sociology at the University of New Hampshire, but she didn't really know what kind of career it would lead to. Then she took an undergraduate statistics course and found she really enjoyed it. She took additional methods courses—survey research and an individual research project course—and really liked those also.

By the time she graduated, LeBlanc knew she wanted a job in social research. She looked online for research positions in marketing, health care, and other areas. She noticed an opening at a university-based research center and thought their work sounded fascinating. As a research assistant, LeBlanc designed survey questions, transcribed focus group audiotapes, programmed web surveys, and managed incoming data. She also conducted interviews, programmed computer-assisted telephone surveys, and helped conduct focus groups.

The knowledge that LeBlanc gained in her methods courses about research designs, statistics, question construction, and survey procedures prepared her well for her position. Her advice to aspiring researchers: Pay attention in your first methods class!

Strengths and Limitations of Social Research

Using social scientific research methods to develop answers to questions about the social world reduces the likelihood of making everyday errors in reasoning. The various projects that we have reviewed in this chapter illustrate this point:

- A clear definition of the population of interest in each study increased the researchers' ability to draw conclusions without overgeneralizing findings to groups to which they did not apply. Selection of a data set based on a broad, representative sample of the population enabled McPherson et al. (2006) to describe social ties throughout the United States rather than among some unknown set of their friends or acquaintances. The researchers' explicit recognition that persons who do not speak English were not included in their data set helps prevent overgeneralization to groups that were not actually studied (McPherson et al. 2006:356).
- The use of surveys in which each respondent was asked the same set of questions reduced the risk of selective or inaccurate observation, as did careful attention to a range of measurement issues (McPherson et al. 2006:355–356).
- The risk of illogical reasoning was reduced by carefully describing each stage of the research, clearly presenting the findings, and carefully testing the bases for cause-and-effect conclusions. For example, Ling and Stald (2010) tested to see whether age or gender, rather than cell phone use, might have increased the tightness of social group ties in Norway.
- Resistance to change was reduced by providing free computers to participants in the Internet health study (Lindsay et al. 2007:100). The publications by all the researchers help other researchers critique and learn from their findings as well as inform the general public.

Nevertheless, I would be less than honest if I implied that we enter the realm of truth and light when we conduct social research or when we rely solely on the best available social research. Research always has some limitations and some flaws (as does any human endeavor), and our findings are always subject to differing interpretations. Social research permits us to see more, to observe with fewer distortions, and to describe more clearly to others what our opinions are based on, but it will not settle all arguments. Others will always have differing opinions, and some of those others will be social scientists who have conducted their own studies and drawn different conclusions.

Although Nie and Erbring (2000) concluded that the use of the Internet diminished social relations, their study at Stanford was soon followed by the UCLA Center for Communication Policy (2001) and by others at the Pew Internet & American Life Project (Boase et al. 2006). These more recent studies also used survey research methods, but their findings suggested that the use of the Internet does *not* diminish social relations.

Psychologist Robert Kraut's early research suggested that Internet use was isolating, but the recent research by Patulny and Seaman (2017) and others indicates more positive effects. To what extent do different conclusions result from differences in research methods, from different perspectives on similar findings, or from rapid changes in the population of Internet users?

It's not easy to answer such questions, so one research study often leads to another, and another, each one improving on previous research or examining a research question from a somewhat different angle. Part of becoming a good social researcher involves learning that we have to evaluate critically each research study and weigh carefully the entire body of research about a research question before coming to a conclusion. And we have to keep an open mind about alternative interpretations and the possibility of new discoveries. The social phenomena we study are often complex, so we must consider this complexity when we choose methods to study social phenomena and when we interpret the results of these studies.

Alternative Research Orientations

In addition to deciding on the type of research they will conduct, social researchers also must choose among several alternative orientations to research.

Quantitative and/or Qualitative Methods

Different research methods provide different perspectives on social phenomena and so have different strengths and weaknesses. The most general distinction is between quantitative and qualitative research methods. Patulny and Seaman (2017) analyzed the number of different types of social contacts; we say that this study used quantitative methods. Numerical data were also used in the descriptive survey about Facebook users and their age, education, and other characteristics (Rainie et al. 2013), as well as in the Lewis et al. (2008) and Ling and Stald (2010) research cited earlier. In contrast, Hampton and Gupta (2008) observed Wi-Fi users in public spaces. Because the researchers recorded their actual observations and did not attempt to quantify what they were studying, we say that Hampton and Gupta used qualitative methods.

Quantitative methods are most often used when the motives for research are explanation, description, or evaluation. Exploration is more often—although by no means always—the motive for using qualitative methods. I highlight several other differences between quantitative and qualitative methods in Chapter 2. Chapters 10 and 11 present qualitative methods in much more detail.

Quantitative methods: Methods such as surveys and experiments that record variation in social life in terms of categories that vary in amount. Data that are treated as quantitative are either numbers or attributes that can be ordered by magnitude.

Qualitative methods: Methods such as participant observation, intensive interviewing, and focus groups that are designed to capture social life as participants experience it rather than in categories predetermined by the researcher. These methods rely on written or spoken words or observations that do not often have a direct numerical interpretation and typically involve exploratory research questions, inductive reasoning, an orientation to social context and human subjectivity, and the meanings attached by participants to events and to their lives.

Important as it is, I don’t want to place too much emphasis on the distinction between quantitative and qualitative orientations or methods. Social scientists often combine these methods to enrich their understanding of the social world. For example, Hampton and Wellman (2000) used surveys to generate counts of community network usage and other behaviors in Netville, but they also observed social interaction and recorded spoken comments to help interpret these behaviors. In this way, qualitative data about social settings can be used to better understand patterns in quantitative data (Campbell and Russo 1999:141).

Philosophical Perspectives

Your preferences for particular research methods will be shaped in part by your general assumptions about how the social world can best be investigated—by your social research philosophy. The scientific approach reflects the belief that there is an objective reality apart from the perceptions of those who observe it. This is the philosophy traditionally associated with natural science and with the belief that scientists must be objective and unbiased to see reality clearly (Weber 1949:72). Positivism asserts that a well-designed test of a specific prediction—for example, the prediction that social ties decrease among those who use the Internet more—can move us closer to understanding actual social processes. Quantitative researchers are often guided by a positivist philosophy.

Postpositivism is a philosophy that is closely related to positivism because it also assumes an external, objective reality, but postpositivists acknowledge the complexity of this reality and the limitations and biases of the scientists who study it (Guba and Lincoln 1994:109–111). For example, postpositivists may worry that researchers who are heavy computer users themselves will be biased in favor of finding positive social effects of computer use. As a result of concerns such as this, postpositivists do not think we can ever be sure that scientific methods allow us to perceive objective reality. Instead, they believe that the goal of science is to achieve intersubjective agreement among scientists about the nature of reality (Wallace 1983:461). We can be more confident in the conclusions of community of social researchers than in those of any individual social scientist (Campbell and Russo 1999:144).

Interpretivism is a research philosophy that emphasizes the importance of understanding subjective meanings people give to reality; unlike positivism and postpositivism, it does not assume that social processes can be identified objectively.

Some qualitative researchers are guided by an interpretivist philosophy. You will learn about different orientations in Chapter 10.

Basic Science or Applied Research

The effort to figure out what the world is like and why it works as it does—academic motivations—is the goal of basic science (Hammersley 2008:50). The Patulny and Seaman (2017) study is a good example.

Social research may also have more immediate, practical concerns. Evaluation research like that conducted by Hampton and Wellman (1999) on the effect of the Internet on community life seeks to determine whether one program or policy has a more desirable impact than another does. This knowledge can then lead to practical changes, such as increasing community members' access to the Internet so that their possibilities for social relations will expand. Evaluation research and other social research motivated by practical concerns are termed applied research.

Whether you think you would prefer a basic or applied orientation in social research, you have lots of company.

Positivism: The belief, shared by most scientists, that there is a reality that exists quite apart from our own perception of it, that it can be understood through observation, and that it follows general laws.

Postpositivism: A philosophical view that modifies the positivist premise of an external, objective reality by recognizing its complexity, the limitations of human observers, and therefore the impossibility of developing more than a partial understanding of reality.

Intersubjective agreement: Agreement between scientists about the nature of reality; often upheld as a more reasonable goal for science than certainty about an objective reality.

Interpretivism: The belief that the subjective meanings people give to their experiences are a key focus for social science research without assuming that social processes can be identified objectively.

Basic science: Research conducted using the scientific method and having the goal of figuring out what the world is like and why it works as it does.

Applied research: Research conducted using the scientific method and addressing immediate, practical concerns, such as determining whether one program or policy has a more desirable impact than another.

The Role of Values

The positivist and postpositivist philosophies consider value considerations to be beyond the scope of science: "An empirical science cannot tell anyone what he should do—but rather what he can do—and under certain circumstances—what he wishes to do" (Weber 1949:54). The idea is that developing valid knowledge about how society *is* organized, or how we live our lives, does not tell us how society *should* be organized or how we *should* live our lives. The determination of empirical facts should be a separate process from the evaluation of these facts as satisfactory or unsatisfactory (Weber 1949:11). The idea is not to ignore value considerations, but to hold them in abeyance during the research project until results are published.

There has always been tension between this "value-free" orientation to social research and a more "value-conscious" or even activist approach. In the 19th century, sociologist Lester Frank Ward (who subsequently became the American Sociological Society's first president) argued that "the real object of science is to benefit man. A science which fails to do this, however agreeable its study, is lifeless" (Ward 1897:xxvii). However, the 1929 American Sociological Society president, William Fielding Ogburn, urged the value-free approach: "Sociology as a science is not interested in making the world a better place to live. . . . Science is interested directly in one thing only, to wit, discovering new knowledge" (Ogburn 1930:300–301). Does one approach make more sense to you?

By the time you finish *Investigating the Social World,* I know you'll have a good understanding of the difference between these orientations, but I can't predict whether you'll decide one is preferable. Maybe you'll conclude that they each have some merit.

Whether you plan to conduct your own research projects, read others' research reports, or just think about and act in the social world, recognizing the strengths and limitations of specific research methods and different approaches to social research will give you greater confidence in your own opinions; improve your ability to evaluate others' opinions; and encourage you to refine your questions, answers, and methods of inquiry.

Conclusions

I hope this first chapter has given you an idea of what to expect from the rest of the book. My aim is to introduce you to social research methods by describing what social scientists have learned about the social world as well as how they have learned it. The substance of social science is inevitably more interesting than its methods, but the methods become more interesting when they're linked to substantive investigations. I have focused attention in this chapter on research about social ties; in subsequent chapters, I introduce research examples from other areas.

Investigating the Social World is organized into four sections. The first section, Foundations for Social Research, includes the introduction in Chapter 1, and then an overview of the research process in Chapter 2 and an introduction to issues in research ethics in Chapter 3. In Chapter 2, I review how social scientists select research questions for investigation, how they orient themselves to those questions with social theories, and how they review related prior research. Most of the chapter focuses on the steps involved in the overall research process and the criteria that researchers use to assess the quality of their answers to the original research questions. Several studies of domestic violence illustrate the research process in Chapter 2. Chapter 3, on research ethics and research proposals, completes the foundation for our study of social research. I emphasize in this chapter and the end-of-chapter exercises the importance of ethical treatment of human subjects in research. I also introduce in this chapter the process of writing research proposals, which I then continue in the end-of-chapter exercises throughout the book. In actual research projects, submission of a research proposal to an Institutional Review Board for the Protection of Human Subjects is often the final step in laying the foundation for a research project.

The second section, Fundamentals of Social Research, presents methods for conceptualization and measurement, sampling, and causation and other elements of research design that must be considered in any social research project. In Chapter 4, I discuss the concepts we use to think about the social world and the measures we use to collect data about those concepts. This chapter begins with the example of research on student substance abuse, but you will find throughout this chapter a range of examples from contemporary research. In Chapter 5, I use research on homelessness to exemplify the issues involved in sampling cases to study. In Chapter 6, I use research on violence to illustrate how to design research to answer such causal research questions as "What causes violence?" I also explain in this chapter the decisions that social researchers must make about two research design issues that affect our ability to draw causal conclusions: (1) whether to use groups or individuals as units of analysis and (2) whether to use a cross-sectional or longitudinal research design.

The third section, Basic Social Research Designs, introduces the three primary methods of data collection and related methods of data analysis. Experimental studies, the subject of Chapter 7, focus attention on testing causal effects and are used often by social psychologists, psychologists, and policy evaluation researchers. Survey research is the most common method of data collection in sociology, so in Chapter 8, I describe the different types of surveys and explain how researchers design survey questions. I highlight in this chapter the ways in which the Internet and cell phones are changing the nature of survey research. The next chapter, on quantitative data analysis, introduces the statistics used to analyze data collected with experimental and survey designs. Chapter 9 is not a substitute for an entire course in statistics, but it provides the basic tools you can use to answer most research questions. To make this chapter realistic, I walk you through an analysis of quantitative data on voting in the 2008 presidential election. You can replicate this analysis with data on the book's study site (if you have access to the SPSS statistical analysis program). You can also learn more about statistics with the SPSS exercises at the end of most chapters and with the study site's tutorials.

Qualitative methods have long been the method of choice in anthropology, but they also have a long tradition in American sociology and have become the favored method of many social researchers around the world. Chapter 10 shows how qualitative techniques can uncover aspects of the social world that we are likely to miss in experiments and surveys and can sometimes result in a different perspective on social processes. Chapter 11 then focuses on the logic and procedures of analyzing qualitative data. In these chapters, you will learn about research on work organizations, psychological distress, gender roles, classroom behavior, and disasters such as Hurricane Katrina.

The fourth section, Complex Social Research Designs, presents research designs that can involve combinations of one or more of the basic research designs. By the time you read Chapter 12, you should be convinced of the value of using different methods to help us understand different aspects of the social world. Chapter 12 takes this basic insight a few steps further by introducing the use of "mixed methods." This increasingly popular approach to research design involves a careful plan for combining qualitative and quantitative methods in a research project. Evaluation research, the subject of Chapter 13, is conducted to identify the impact of social programs or to clarify social processes involving such programs. Evaluation research often uses experimental methods, but survey research and qualitative methods can also be helpful in evaluation research projects. Chapter 14 reviews the methods of secondary data analysis and the related approach that has come to be known as "Big Data." In this chapter, you will learn how to obtain previously collected data and to investigate important social issues such as poverty dynamics.

Historical and comparative methods, the subject of Chapter 15, may involve either quantitative or qualitative methods that are used to compare societies and groups at one point in time and to analyze their development over time. We will see how these different approaches have been used to learn about political change in transitional societies. I also explain the method of content analysis in this chapter; it can be used in historical and comparative research and provides another way to investigate social processes in an unobtrusive way.

Plan to read Chapter 16 carefully. Our research efforts are only as good as the attention given to our research reports, so my primary focus in this chapter is on writing research reports. I also present means for enhancing graphic displays to communicate quantitative results more effectively in research reports. In addition, I introduce meta-analysis—a statistical technique for assessing many research studies about a particular research question. By the end of the chapter, you should have a broader perspective on how research methods can improve understanding of the social world (as well as an appreciation for how much remains to be done).

Each chapter ends with several helpful learning tools. Lists of key terms and chapter highlights will help you review the ideas that have been discussed. Discussion questions and practice exercises will help you apply and deepen your knowledge. Special exercises guide you in developing your first research proposal, finding information on the Internet, grappling with ethical dilemmas, and conducting statistical analyses.

The study site for this book on the SAGE website provides interactive exercises and quizzes for reviewing key concepts, as well as research articles to review, websites to visit, data to analyze, and short lectures to hear. Check it out at edge.sagepub.com/schutt9e.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Applied research 20
Basic science 20
Descriptive research 14
Evaluation research 17
Explanatory research 16
Exploratory research 15
Illogical reasoning 9
Inaccurate observation 7
Interpretivism 20
Intersubjective agreement 20
Overgeneralization 8
Positivism 20
Postpositivism 20
Pseudoscience 12
Qualitative methods 19
Quantitative methods 19
Resistance to change 9
Science 10
Selective observation 7
Social science 12

Highlights

- Social research differs from the ordinary process of thinking about our experiences by focusing on broader questions that involve people outside our immediate experience and issues about why things happen, and by using systematic research methods to answer those questions. Four common errors in reasoning are (1) selective or inaccurate observation, (2) overgeneralization, (3) illogical reasoning, and (4) resistance to change. These errors result from the complexity of the social world, subjective processes that affect the reasoning of researchers and those they study, researchers' self-interestedness, and unquestioning acceptance of tradition or of those in positions of authority.
- Social science is the use of logical, systematic, documented methods to investigate individuals, societies, and social processes, as well as the knowledge produced by these investigations.
- Social research cannot resolve value questions or provide permanent, universally accepted answers.
- Social research can be motivated by policy guidance and program management needs, academic concerns, and personal or charitable impulses.
- Social research can be descriptive, exploratory, explanatory, or evaluative—or some combination of these.
- Quantitative and qualitative methods structure research in different ways and are differentially appropriate for diverse research situations. They may be combined in research projects.
- Positivism and postpositivism are research philosophies that emphasize the goal of understanding the real world; these philosophies guide most quantitative researchers.
- Interpretivism is a research philosophy that emphasizes an understanding of the meaning people attach to their experiences; it guides many qualitative researchers.
- Basic science research focuses on expanding knowledge and providing results to other researchers.

Applied research seeks to have an impact on social practice and to share results with a wide audience.

Discussion Questions

1. Select a social issue that interests you, such as Internet use or crime. List at least four of your beliefs about this phenomenon. Try to identify the sources of each of these beliefs.
2. Does the academic motivation to do the best possible job of understanding how the social world works conflict with policy or personal motivations? How could personal experiences with social isolation or with Internet use shape research motivations? In what ways might the goal of influencing policy about social relations shape a researcher's approach to this issue?
3. Pick a contemporary social issue of interest to you. Describe different approaches to research on this issue that would involve descriptive, exploratory, explanatory, and evaluative approaches.
4. Review the strengths of social research. How convinced are you about each of them at this point?
5. Review each of the research alternatives. Do you find yourself more attracted to a quantitative or a qualitative approach? To a positivist, postpositivist, or interpretivist philosophy? To doing research to contribute to basic knowledge or to shape social policy? What do you think about value freedom as a standard for science?

Practice Exercises

1. Read the abstracts (initial summaries) of five articles available in the “article review matrix” on the study site. On the basis of the abstract only, classify each research project represented in the articles as primarily descriptive, exploratory, explanatory, or evaluative. Note any indications that the research focused on other types of research questions.
2. Find a report of social science research in an article in a daily newspaper. What are the motives for the research? How much information is provided about the research design? What were the major findings? What additional evidence would you like to see in the article to increase your confidence in the research conclusions?
3. Review “Types of Research” from the “Interactive Exercises” link on the study site. To use these lessons, choose one of the four “Types of Research” exercises from the opening menu. About 10 questions are presented in each version of the lesson. After reading each question, choose one answer from the list presented. The program will evaluate your answers. If an answer is correct, the program will explain why you were right and go on to the next question. If you have made an error, the program will explain the error to you and give you another chance to respond.
4. Now, select a journal article from edge.sagepub.com/schutt9e and read its abstract. Identify the type of research (descriptive, exploratory, explanatory, or evaluation) that appeared to be used. Now scan the article and decide whether the approach was quantitative or qualitative (or both) and whether it included any discussion of policy implications.

Ethics Questions

Throughout the book, we will discuss the ethical challenges that arise in social research. At the end of each chapter, you are asked to consider some questions about ethical issues related to that chapter's focus. I introduce this critical topic formally in Chapter 3, but we will begin here with some questions for you to ponder.

1. The chapter refers to research on social isolation. What would *you* do if you were interviewing elderly persons in the community and found that one was very isolated and depressed or even suicidal, apparently as a result of his or her isolation? Do you believe that social researchers have an obligation to take action in a situation like this? What if you discovered a similar problem with a child? What guidelines would you suggest for researchers?
2. Would you encourage social researchers to announce their findings in press conferences about topics such as the impact of the Internet on social ties, and to encourage relevant agencies to adopt policies aimed to lessen social isolation? Are there any advantages to studying research questions only to contribute to academic knowledge? Do you think there is a fundamental conflict between academic and policy motivations? Do social researchers have an ethical obligation to recommend policies that their research suggests would help other people?

Web Exercises

1. The research on social ties by McPherson and his colleagues was publicized in a *Washington Post* article that also included comments by other sociologists ("Social Isolation Growing in U.S., Study Says," by Shankar Vedantam). Read the article at www.washingtonpost.com/wp-dyn/content/article/2006/06/22/AR2006062201763_pf.html and continue the commentary. Do your own experiences suggest that there is a problem with social ties in your community? Does it seem, as Wellman suggests in the *Washington Post* article, that a larger number of social ties can make up for the decline in intimate social ties that McPherson found?
2. Scan one of the publications about the Internet and society at the Berkman Klein Center for Internet & Society website, https://cyber.harvard.edu/. Describe one of the projects discussed: its goals, methods, and major findings. What do the researchers conclude about the impact of the Internet on social life in the United States? Next, repeat this process with a report from the Pew Research Center Internet & Technology website at www.pewinternet.org, or with the Digital Future report from the University of Southern California's Center for the Digital Future site, www.digitalcenter.org. What aspects of the methods, questions, or findings might explain differences in their conclusions? Do you think the researchers approached their studies with different perspectives at the outset? If so, what might these perspectives have been?

Video Interview Questions

Listen to the researcher interview for Chapter 1 at edge.sagepub.com/schutt9e.

1. What are the benefits to breaking down questions in a text-based interview structure?
2. As Janet Salmons mentions, one can enhance his or her research by deciding carefully on the various kinds of technology to be used. What are some of the considerations that Salmons suggests to help decide whether to use text-based interviews or video conference calls?

SPSS Exercises

The SPSS exercises at the end of each chapter focus on support for the death penalty. A portion of the 2016 GSS survey data as well as the complete GSS2016 data set are available on the study site. You will need to use one of these files to carry out these exercises. If you are able to use the complete version of SPSS (perhaps in your university's computer lab), download the GSS2016 or GSS2016x file. If you are using the student version of SPSS, download the GSS2016x_reduced file. You will begin your empirical investigation by thinking a bit about the topic and the data you have available for study.

1. What personal motivation might you have for studying support for the death penalty? What might motivate other people to conduct research on this topic? What policy and academic motives might be important?
2. After you download one of the GSS2016 files and save it in a directory, open the GSS2016, GSS2016x, or GSS2016x_reduced file. In the SPSS menu, click on File, then on Open and Data, and then on the name of the data file in the directory where it is saved. How many respondents are there in this subset of the complete GSS file? (Scroll down to the bottom of the data set in Data View.) How many variables were measured? (Scroll down to the bottom of the Variable View in SPSS.)
3. What would you estimate as the level of support for capital punishment in the United States in 2016? Now for your first real research experience in this text: Describe the distribution of support for capital punishment. Obtaining the relevant data is as simple as "a, b, c, d, e."
   a. Click on Graphs.
   b. Click on Legacy Dialogs > Bar.
   c. Select "Simple" and "Summaries for groups of cases" under Data in Chart Area > Define.
   d. Place the CAPPUN variable in the box below "Category Axis:" and select "% of cases" under "Bar Represent."
   e. Click OK.

Now describe the distribution of support for capital punishment. What percentage of the population supported capital punishment in the United States in 2016?

Developing a Research Proposal

Will you develop a research proposal in this course? If so, you should begin to consider your alternatives:

1. What topic would you focus on, if you could design a social research project without any concern for costs? What are your motives for studying this topic?
2. Develop four questions that you might investigate about the topic you just selected. Each question should reflect a different research motive: description, exploration, explanation, or evaluation. Be specific.
3. Which question most interests you? Would you prefer to attempt to answer that question with quantitative or qualitative methods? Do you seek to contribute to basic science or to applied research?

Chapter 2 The Process and Problems of Social Research

- Research That Matters, Questions That Count
- Social Research Questions
  - Identifying Social Research Questions
  - Refining Social Research Questions
  - Evaluating Social Research Questions
    - *Feasibility*
    - *Social Importance*
    - *Scientific Relevance*
- Social Theories
  - Scientific Paradigms
- Social Research Foundations
  - Searching the Literature
  - Reviewing Research
    - *Single-Article Reviews: Formal and Informal Deterrents to Domestic Violence*
    - *Integrated Literature Reviews: When Does Arrest Matter?*
    - *Systematic Literature Reviews: Second Responder Programs and Repeat Family Abuse Incidents*
  - Searching the Web
- Social Research Strategies
- Research in the News: Control and Fear: What Mass Killings and Domestic Violence Have in Common
  - Explanatory Research
    - *Deductive Research*
    - *Domestic Violence and the Research Circle*
    - *Inductive Research*
  - Exploratory Research
    - *Battered Women's Help Seeking*
  - Descriptive Research
- Careers and Research
- Social Research Organizations
- Social Research Standards
  - Measurement Validity
  - Generalizability
  - Causal Validity
  - Authenticity
- Conclusions

Research That Matters, Questions That Count

The High Point Police Department (HPPD) in High Point, North Carolina, developed the Offender Focused Domestic Violence Initiative (OFDVI) in an attempt to reduce repeat offenses by known offenders. The OFDVI concentrated both law enforcement and social service resources on repeat offenders by notifying them of swift, certain, and potentially severe consequences if they reoffend and offered resource assistance, with responses varying with prior record. Did this new approach reduce reoffending? Researchers from the University of North Carolina at Greensboro worked with local and federal law enforcement and social service representatives to develop the intervention and an evaluation strategy. The HPPD changed its procedures to distinguish interpersonal domestic violence (IPDV) calls and arrests. When Stacy Sechrist and John Weil analyzed the data collected before and after OFDVI was implemented, they found that IPDV arrest incidents had declined by 20%. Declines also occurred in calls to police and in the rate of victim injury, but not in the rate of offender recidivism.

1. Can you think of other possible explanations for the changes in IPDV calls, arrests, and injuries?
2. Why do you think the notification intervention might have reduced arrests but not reoffending?

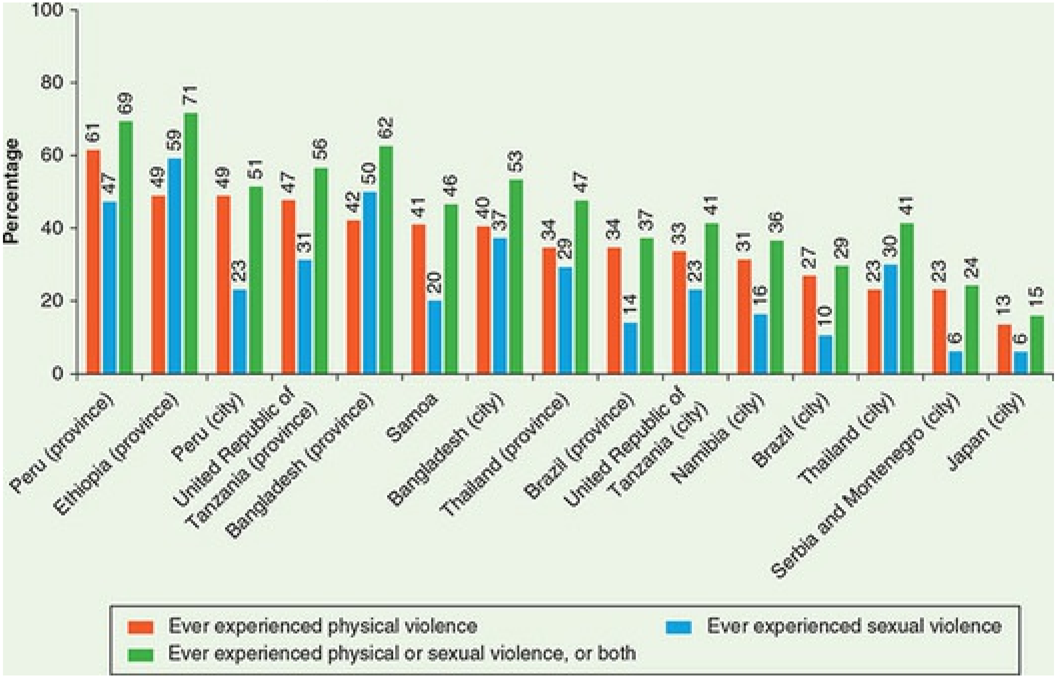
In this chapter, you will learn about methods used to study the response to domestic violence and the conclusions from some of this research. By the end of the chapter, you will have a much firmer basis for answering the questions I have posed. You can learn more about the issues by reading the complete 2017 *Violence Against Women* article by Stacy Sechrist and John Weil at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 2 at edge.sagepub.com/schutt9e.

Sechrist, Stacy M. and John D. Weil. 2017. "Assessing the Impact of a Focused Deterrence Strategy to Combat Intimate Partner Domestic Violence." *Violence Against Women.* Online first.

Domestic violence is a major problem in countries around the world. An international survey by the World Health Organization (WHO) of 24,000 women in 10 countries estimated lifetime physical or sexual abuse ranging from a low of 15% in Japan to a high of 71% in rural Ethiopia (WHO 2005:6) (see Exhibit 2.1). In a 2010 U.S. survey of 16,507 men and women sponsored by the Department of Justice and the Centers for Disease Control and Prevention, 35.6% of women and 28.5% of men said they had experienced rape, physical violence, or stalking by an intimate partner at some time in their lives (Black et al. 2011), although rates of intimate partner violence (IPV) had declined by 64% in the United States since 1994 (Catalano 2012).

What can be done to reduce this problem? Long before the OFDVI program in High Point, the Police Foundation and the Minneapolis Police Department began research to answer this question with an experiment to determine whether arresting accused spouse abusers on the spot would deter repeat incidents. The study's results, which were publicized widely, indicated that arrests did have a deterrent effect. Partly because of this, the percentage of urban police departments that made arrest the preferred response to complaints of domestic violence rose from 10% in 1984 to 90% in 1988 (Sherman 1992:14). Researchers in six other locations then conducted similar experiments to determine whether changing the location or other research procedures would result in different outcomes (Sherman 1992; Sherman and Berk 1984). Many other studies have been conducted since these original seven experiments to better understand issues involved in responding to domestic violence, including a recent 23-year follow-up to determine the long-term effects of the police intervention in the Milwaukee replication of the Minneapolis experiment (Sherman and Harris 2013). The Minneapolis Domestic Violence Experiment, the additional research inspired by it, and the controversies arising from it will provide good examples for our systematic overview of the social research process.

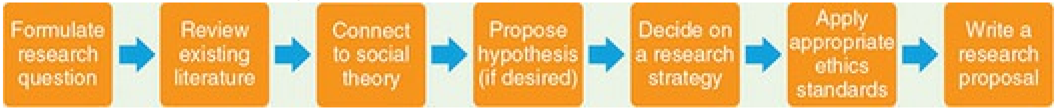
Exhibit 2.1 International Prevalence of Lifetime Physical and Sexual Violence by an Intimate Partner, Among Ever-Partnered Women, by Site



Source: World Health Organization. 2005. Multi-country Study on Women's Health and Domestic Violence Against Women: Summary Report.

In this chapter, we shift from examining the *why* of social research to an overview of the *how*—the focus of the rest of the book. We first consider how to develop a question for social research, then how to review the existing literature about this question while connecting the question to social theory and, in many studies, formulating specific testable hypotheses (see Exhibit 2.2). We then discuss different social research strategies and standards for social research as a prelude to the details about these stages in subsequent chapters. You will find more details in Appendixes A and B about reviewing the literature. I will use the Minneapolis experiment and the related research to illustrate the different research strategies and some of the related techniques. The chapter also expands on the role of social theories in developing research questions and guiding research decisions. By the chapter's end, you should be ready to formulate a research question, critique previous studies that addressed this question, and design a general strategy for answering the question. In the next chapter, you will learn how to review ethical issues and write a research proposal.

Exhibit 2.2 Launching a Research Project


Formulate research question
Review existing literature
Connect to social theory
Propose hypothesis (if desired)
Decide on a research strategy
Apply appropriate ethics standards
Write a research proposal

Social Research Questions

A social research question is a question about the social world that one seeks to answer through the collection and analysis of firsthand, verifiable, empirical data. It is not a question about who did what to whom, but a question about people in groups, about general social processes, or about tendencies in community change such as the following: What distinguishes Internet users from other persons? How do people react to social isolation? Does community policing reduce the crime rate? What influences the likelihood of spouse abuse? So many research questions are possible that it is more of a challenge to specify what does *not* qualify as a social research question than to specify what does.

But that doesn't mean it is easy to specify a research question. Actually, formulating a good research question can be surprisingly difficult. We can break the process into three stages: (1) identifying one or more questions for study, (2) refining the questions, and then (3) evaluating the questions.

Social research question: A question about the social world that is answered through the collection and analysis of firsthand, verifiable, empirical data.

Identifying Social Research Questions

Social research questions may emerge from your own experience—from your "personal troubles," as C. Wright Mills (1959) put it. One experience might be membership in a church, another could be victimization by crime, and yet another might be moving from a dorm to a sorority house. You may find yourself asking a question such as "In what ways do people tend to benefit from church membership?" "Does victimization change a person's trust in others?" or "How do initiation procedures influence group commitment?" What other possible research questions can you develop based on your own experiences in the social world? You should focus on a research question that interests you and others (Firebaugh 2008:2).

The research literature is often the best source for research questions. Every article or book will bring new questions to mind. Even if you're not feeling too creative when you read the literature, most research articles highlight unresolved issues and end with suggestions for additional research. For example, Richard A. Berk, Alec Campbell, Ruth Klap, and Bruce Western (1992) concluded an article on four of the replications of the Minneapolis experiment on police responses to spouse abuse by suggesting, "Deterrence may be effective for a substantial segment of the offender population. . . . However, the underlying mechanisms remain obscure" (p. 706). A new study could focus on these mechanisms: Why does the arrest of offenders deter some of them from future criminal acts? Is it just incapacitation due to court processes and incarceration? Fear of punishment? Do their attitudes about abuse change? Any research article in a journal is likely to suggest other unresolved issues about the research question studied.

Many social scientists find the source of their research questions in social theory. Some researchers spend much of their careers conducting research intended to refine an answer to one research question that is critical for a particular social theory. For example, one theory explains social deviance as resulting from labels that people attach to some disordered behaviors. A researcher could ask whether this "labeling theory" can explain how spouse abusers react to being arrested.

Finally, some research questions have very pragmatic sources. You may focus on a research question someone else posed because it seems to be to your advantage to do so. Some social scientists conduct research on specific questions posed by a funding source in what is termed an *RFP,* a request for proposals. (Sometimes the acronym RFA is used, meaning request for applications.) The six projects to test the conclusions of the Minneapolis Domestic Violence Experiment were developed in response to such a call for proposals from the National Institute of Justice. Or you may learn that the social workers in the homeless shelter where you volunteer need help with a survey to learn about client needs, which becomes the basis for another research question.

Refining Social Research Questions

It is even more challenging to focus on a problem of manageable size than it is to come up with an interesting question for research. We are often interested in much more than we can reasonably investigate with limited time and resources. In addition, researchers may worry about staking a research project (and thereby a grant or a grade) on a single problem, and so they may address several research questions at once. Also, it might seem risky to focus on a research question that may lead to results discrepant with our own cherished assumptions about the social world. The prospective commitment of time and effort for some research questions may seem overwhelming, resulting in a certain degree of paralysis.

The best way to avoid these problems is to develop the research question gradually. Don't keep hoping that the perfect research question will just spring forth from your pen. Instead, develop a list of possible research questions as you go along. At the appropriate time, look through this list for the research questions that appear more than once. Narrow your list to the most interesting, most workable candidates. Repeat this process as long as it helps improve your research question.

Evaluating Social Research Questions

In the third stage of selecting a research question, you evaluate the best candidate against the criteria for good social research questions: feasibility, given the time and resources available; social importance; and scientific relevance (King, Keohane, and Verba 1994).

Feasibility

You must be able to conduct any study within the time and resources available. If time is short, questions that involve long-term change may not be feasible. Another issue is to what people or groups we can expect to gain access. For example, observing social interaction in corporate boardrooms may be taboo. Next, you must consider whether you will have any additional resources, such as research funds or other researchers with whom to collaborate. Remember that there are severe limits on what one person can accomplish. However, you may be able to piggyback your research onto a larger research project. You also must consider the constraints faced because of your schedule, other commitments, and skill level, not to mention what is ethically defensible.

The Minneapolis Domestic Violence Experiment shows how ambitious a social research question can be when a team of seasoned researchers secures the backing of influential groups. The project required hundreds of thousands of dollars, the collaboration of many social scientists and criminal justice personnel, and the volunteer efforts of 41 Minneapolis police officers. Of course, for this reason, the Sherman and Berk (1984) question would not be feasible for a student project. You might instead ask the question “Do students think punishment deters spouse abuse?” Or perhaps you could work out an arrangement with a local police department to study the question “How satisfied are police officers with their treatment of domestic violence cases?”

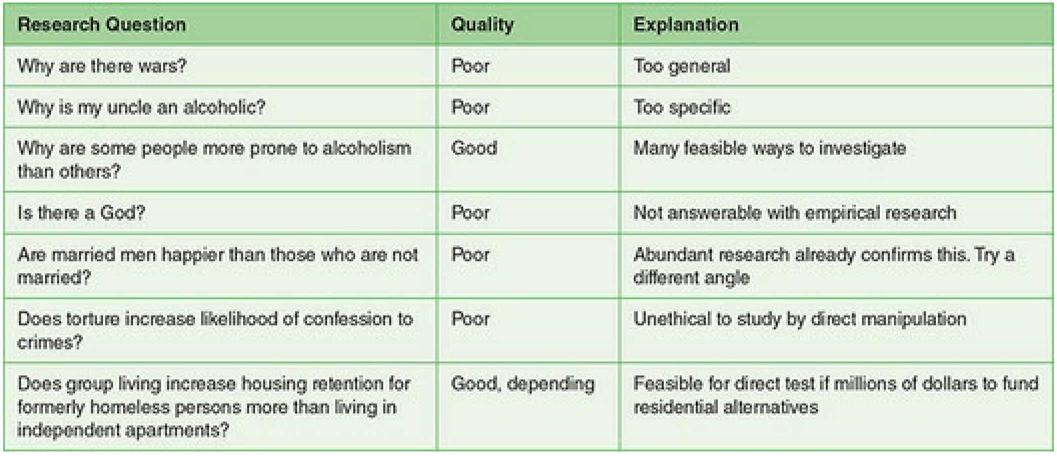
Social Importance

Social research is not a simple undertaking, so it’s hard to justify the expenditure of effort and resources unless you focus on a substantive area that is important. Besides, you need to feel motivated to carry out the study. Nonetheless, “importance” is relative, so for a class assignment, student reactions to dormitory rules or something similar might be important enough.

For most research undertakings, you should consider whether the research question is important to other people. Will an answer to the research question make a difference for society or for social relations? Again, the Minneapolis Domestic Violence Experiment is an exemplary case. But the social sciences are not wanting for important research questions. The August 2013 issue of the *American Sociological Review*—the journal that published the first academic article on the Minneapolis experiment—contained articles on environment, urbanization, leadership, migration processes, social class and lifestyle, and global social change. All these articles addressed research questions about important social issues, and all raised new questions for additional research.

The examples of good and poor research questions in Exhibit 2.3 give you an idea of how to apply these criteria.

Exhibit 2.3 Evaluating Research Questions



Research Question	Quality	Explanation
Why are there wars?	Poor	Too general
Why is my uncle an alcoholic?	Poor	Too specific
Why are some people more prone to alcoholism than others?	Good	Many feasible ways to investigate
Is there a God?	Poor	Not answerable with empirical research
Are married men happier than those who are not married?	Poor	Abundant research already confirms this. Try a different angle
Does torture increase likelihood of confession to crimes?	Poor	Unethical to study by direct manipulation
Does group living increase housing retention for formerly homeless persons more than living in independent apartments?	Good, depending	Feasible for direct test if millions of dollars to fund residential alternatives

Source: Firebaugh, Glenn. 2008. Seven Rules for Social Research. Princeton, NJ: Princeton University Press.

Scientific Relevance

Every research question should be grounded in the social science literature. Whether you formulate a research question because you have been stimulated by an academic article or because you want to investigate a current social problem, you should first turn to the social science literature to find out what has already been learned about this question. You can be sure that some prior study is relevant to almost any research question you can think of.

The Minneapolis experiment was built on a substantial body of contradictory theorizing about the impact of punishment on criminality (Sherman and Berk 1984). Two popular theories that make contradictory predictions about the impact of punishment are deterrence theory and labeling theory. Deterrence theory predicted that arrest would deter individuals from repeat offenses; labeling theory predicted that arrest would make repeat offenses more likely (both theories are described in more detail later in this chapter). Only one prior experimental study of this issue had been conducted with juveniles, and studies among adults had yielded inconsistent findings. Clearly, the Minneapolis researchers had good reason for another study. Any new research question should be connected in this way to past research.

What if there is already a lot of convincing research published about your proposed research question? Can you focus the question on a new population or setting? Is early intervention with social services more effective in reducing violence among juveniles than among adults? Do men like using Facebook as much as women do? You get the idea.

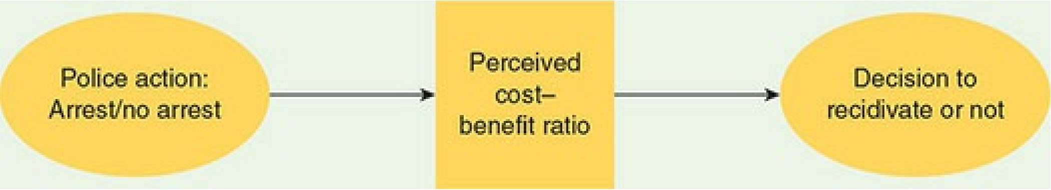
Social Theories

Neither domestic violence nor police policies exist in a vacuum, set apart from the rest of the social world. We can understand the particular behaviors and orientations better if we consider how they reflect broader social patterns. Do abusive men keep their wives in positions of subservience? Are community members law abiding? Our answers to general questions such as these will help shape the research questions that we ask and the methods that we use.

Although everyone has general notions about "how things work," "what people are like," and so on, social scientists draw on more formal sets of general ideas—social theories—to guide their research (Collins 1994). A theory is a logically interrelated set of propositions that helps us make sense of many interrelated phenomena and predict behavior or attitudes that are likely to occur when certain conditions are met. Theory helps social scientists decide which questions are important to ask about the social world and which are just trivial pursuits. Theory focuses a spotlight on the particular features of the social world where we should look to get answers for these questions, how these features are related to each other, and what features can be ignored. Building and evaluating theory is therefore one of the most important objectives of social science.

Theory: A logically interrelated set of propositions about empirical reality.

Exhibit 2.4 Rational Choice Theory Prediction


Police action: Arrest/no arrest
Perceived cost–benefit ratio
Decision to recidivate or not

Lawrence Sherman and Richard Berk's (1984) domestic violence experiment tested predictions from a type of deterrence theory, as I have already mentioned. Deterrence theory is itself based on a broader perspective that is termed rational choice theory. Rational choice theory assumes that people's behavior is shaped by practical cost–benefit calculations (Coleman 1990:14). *Specific deterrence theory* applies rational choice theory to crime and punishment (Lempert and Sanders 1986:86–87). It states that arresting spouse abusers will lessen their likelihood of reoffending by increasing the costs of reoffending. Crime "doesn't pay" (as much) for these people (see Exhibit 2.4). The High Point OFDVI was also based on rational choice theory—let repeat offenders know that the policy has changed and that they will be certain to face severe punishment if they reoffend.

Do these concepts interest you? Do these propositions strike you as reasonable ones? If so, you might join a long list of researchers who have attempted to test, extend, and modify various aspects of rational choice theory.

Raymond Paternoster and his colleagues (1997) concluded that rational choice theory—in particular, specific deterrence theory—did not provide an adequate framework for explaining how citizens respond to arrest. Paternoster et al. turned to procedural justice theory for a very different prediction. Procedural justice theory predicts that people will obey the law from a sense of obligation that flows from seeing legal authorities as moral and legitimate (Tyler 1990). From this perspective, individuals who are arrested will be less likely to reoffend if they are treated fairly, irrespective of the outcome of their case, because fair treatment will enhance their view of legal authorities as moral and legitimate. Procedural justice theory expands our view of the punishment process by focusing attention on *how* authorities treat people rather than just on *what* decisions the authorities make about people.

Some sociologists attempt to understand the social world by looking inward, at the meaning people attach to their interactions. These researchers focus on the symbolic nature of social interaction—how social interaction conveys meaning and promotes socialization. Herbert Blumer developed these ideas into symbolic interaction theory (Turner, Beeghley, and Powers 1995:460).

Labeling theory uses a symbolic interactionist approach to explain deviance as an "offender's" reaction to the application of rules and sanctions (Becker 1963:9; Scull 1988:678). Sherman and Berk (1984) recognized that a labeling process might influence offenders' responses to arrest in domestic violence cases. Once the offender is labeled as a deviant by undergoing arrest, other people treat the offender as deviant, and he or she is then more likely to act in a way that is consistent with the label *deviant.* Ironically, the act of punishment stimulates more of the very behavior that it was intended to eliminate. This theory suggests that persons arrested for domestic assault are more likely to reoffend than are those who are not punished, which is the reverse of the deterrence theory prediction.

Do you find yourself thinking of some interesting research foci when you read about this labeling theory of deviance? If so, consider developing your knowledge of symbolic interaction theory and use it as a guide in your research.

Conflict theory focuses on basic conflicts between different social groups in society and how groups attempt to exercise domination to their own benefit (Collins 1994:47). The theory has its origins in Karl Marx and Friedrich Engels's (1961:13–16) focus on social classes as the key groupings in society and their belief that conflict between social classes was not only the norm but also the "engine" of social change.

Rational choice theory: A social theory that explains individual action with the principle that actors choose actions that maximize their gains from taking that action.

Procedural justice theory: A social theory that predicts that people will obey the law from a sense of obligation that flows from seeing legal authorities as moral and legitimate.

Symbolic interaction theory: A social theory that focuses on the symbolic nature of social interaction—how social interaction conveys meaning and promotes socialization.

Conflict theory: A social theory that identifies conflict between social groups as the primary force in society; understanding the bases and consequences of the conflict is key to understanding social processes.

Although different versions of conflict theory emphasize different bases for conflict, they focus attention on the conflicting interests of groups rather than on the individuals' concerns with maximizing their self-interest. As applied to crime, conflict theory suggests that laws and the criminal justice system are tools of the upper classes to maintain their dominance over lower classes.

Do these concepts strike a responsive chord with you? Can you think of instances in which propositions of conflict theory might help explain social change?

French social theorist Émile Durkheim used a very different theory, functionalism, to explain crime and other forms of deviance in society. Writing during the period of rapid social change in Europe at the dawn of the 20th century, Durkheim (1893/2014) was concerned with the strength of social bonds in society. He posited that traditional social bonds based on similarity between people were being replaced by social bonds based on interdependence between people performing different social roles. For example, urban dwellers needed farmers to grow their food, truckers to bring the crops to market, merchants to arrange the sale of the crops, butchers to prepare meat, cobblers to make shoes, and so forth. Durkheim (1893/2014) termed social bonds based in this way on interdependence as *organic solidarity* (bringing to mind the interdependence of different organs in the body). Crime is explained by functionalists as occurring because it is functional for society to delimit the boundaries around acceptable behavior.

As a social researcher, you may work with one of these theories, seeking to extend it, challenge it, or specify it. You may test alternative implications of the different theories against each other. If you're feeling ambitious, you may even seek to combine some aspects of the different perspectives. Maybe you'll come up with a different theoretical perspective altogether. Or you may find that you lose sight of the larger picture in the midst of a research project; after all, it is easier to focus on accumulating particular findings rather than considering how those findings fit into a more general understanding of the social world. But you'll find that in any area of research, developing an understanding of relevant theories will help you ask important questions, consider reasonable alternatives, and choose appropriate research procedures.

Scientific Paradigms

Scientific paradigms are sets of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings. In his famous book on the history of science, *The Structure of Scientific Revolutions,* Thomas S. Kuhn (1970) argued that most of the time one scientific paradigm is accepted as the prevailing wisdom in a field and that scientists test ideas that make sense within that paradigm. They are conducting what Kuhn called normal science. Only after a large body of contrary evidence accumulates might there be a rapid shift to a new paradigm (Hammersley 2008:46).

Some people refer to conflict theory, functionalist theory, and symbolic interaction theory as alternative paradigms, although this stretches the meaning of *paradigm* a bit. In any case, many social scientists draw on both of these perspectives in their research and so reject the notion that these are truly incommensurable paradigms. We also should be sensitive to the insights that can be provided by examining social phenomena from both perspectives.

Functionalism: A social theory that explains social patterns in terms of their consequences for society as a whole and that emphasizes the interdependence of social institutions and their common interest in maintaining the social order.

Scientific paradigm: A set of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings.

Normal science: The gradual, incremental research conducted by scientists within the prevailing scientific paradigm.

Social Research Foundations

How do you find prior research and theorizing on questions of interest? You may already know some of the relevant material from prior coursework or your independent reading, but that won't be enough. When you begin to focus on a research question, you need to find reports of previous investigations that sought to answer the same research question that you want to answer, not just those that were about a similar topic. If there have been no prior studies of exactly the same research question on which you want to focus, you should seek reports from investigations of very similar research questions. Once you have located reports from prior research similar to the research that you want to conduct, you may expand your search to include investigations about related topics or studies that used similar methods.

It's not possible to overemphasize the importance of searching and reviewing related literature as you begin to investigate a social research question. Consider some of the issues you can address if you use a literature review to refine your research focus:

- Has a consensus already emerged in response to this research question? If not, what are the differences in findings and what remains to be investigated?
- What has not been acknowledged in prior publications? Have methods been fully described?
- What theories have been applied in this research and which seem most consistent with the evidence?
- Which methods have been used in prior investigations? Have any difficulties emerged in their application?

If you keep issues like these in mind as you review the literature, you will create a solid foundation for your own research.

Although it's most important when you're starting out, reviewing the literature is also important at later stages of the research process. Throughout a research project, you will uncover new issues and encounter unexpected problems; at each of these times, you should search the literature to locate prior research on these issues and to learn how others responded to similar problems. Published research that you ignored when you were seeking other research on domestic violence might become very relevant when you have to decide which questions to ask people about their attitudes toward police and other authorities.

Searching the Literature

Conducting a thorough search of the research literature and then reviewing critically what you have found lays an essential foundation for any research project. Fortunately, much of this information can be identified online, without leaving your desktop, and an increasing number of published journal articles can be downloaded directly onto your own computer (depending on your particular access privileges). But just because there's a lot available online doesn't mean that you need to find it *all.* Keep in mind that your goal is to find relevant reports of prior research investigations. The type of reports you should focus on are those that have been screened for quality through critique by other social scientists before publication. Scholarly journals, or *refereed journals* that publish *peer-reviewed articles,* manage this review process. Most often, editors of refereed journals send articles that authors submit to three or more other social scientists for anonymous review. Based on the reviewers' comments, the journal editor then decides whether to accept or reject the article, or to invite the author to "revise and resubmit." This process results in the rejection of most articles (top journals such as the *American Sociological Review* or the *American Journal of Sociology* may reject about 90% of the articles submitted), and those that are ultimately accepted for publication typically have to be revised and resubmitted first. This helps ensure a much higher-quality standard, although journals vary in the rigor of their review standards, and, of course, different reviewers may be impressed by different types of articles; you thus always have to make your own judgment about article quality.

Newspaper and magazine articles may raise important issues or summarize social science research investigations, but they are not an acceptable source for understanding the research literature. The web offers much useful material, including research reports from government and other sources, sites that describe social programs, and even indexes of the published research literature. You may find copies of particular rating scales, reports from research in progress, papers that have been presented at professional conferences, and online discussions of related topics. Web search engines will also find academic journal articles that you can access directly online, although usually for a fee. Most of the published research literature will be available to you online only if you go through the website of your college or university library. The library pays a fee to companies that provide online journals so that you can retrieve this information without paying anything extra yourself. Of course, no library can afford to pay for every journal, so if you can't find a particular issue of a particular journal that you need online, you will have to order the article that you need through interlibrary loan or, if the hard copy of the journal is available, walk over to your library to read it.

As with any part of the research process, your method for searching the literature will affect the quality of your results. Your search method should include the following steps:

1. *Specify your research question.* Your research question should be neither so broad that hundreds of articles are judged relevant nor so narrow that you miss important literature. “Is informal social control effective?” is probably too broad. “Does informal social control reduce rates of burglary in my town?” is probably too narrow. “Is informal social control more effective in reducing crime rates than policing?” provides about the right level of specificity.
2. *Identify appropriate bibliographic databases to search.* Sociological Abstracts or SocINDEX may meet many of your needs, but if you are studying a question about social factors in illness, you should also search in MEDLINE, the database for searching the medical literature. If your focus is about psychological issues, you’ll also want to include a search in the online Psychological Abstracts database, PsycINFO, or the version that also contains the full text of articles, PsycARTICLES. Search Criminal Justice Abstracts if your topic is in the area of criminology or criminal justice, or EconLit, if your topic might be addressed in the economic literature. It will save you a lot of time if you ask a librarian to teach you the best techniques for retrieving the most relevant articles to answer your questions.

   To find articles that refer to a previous publication, such as Sherman and Berk’s study of the police response to domestic violence, the Social Science Citation Index (SSCI) will be helpful. SSCI is an extremely useful tool for tracing the cumulative research in an area across the social sciences. SSCI has a unique “citation searching” feature that allows you to look up articles or books, see who else has cited them in their own work, and find out which articles and books have had the biggest impact in a field.
3. *Create a tentative list of search terms.* List the parts and subparts of your research question and any related issues that you think are important: “informal social control,” “policing,” “influences on crime rates,” and perhaps “community cohesion and crime.” It might help to start with one key article that has focused on your research question, identify the concepts in it, and search those concepts. You can then expand the search with more concepts in the articles you locate (Booth, Sutton, and Papaioannou 2016:115).
4. *Narrow your search.* The sheer number of references you find can be a problem. For example, searching for “social capital” in July 2017 resulted in 7,675 citations in SocINDEX. Depending on the database you are working with and the purposes of your search, you may want to limit your search to English-language publications, to journal articles rather than conference papers or dissertations (both of which are more difficult to acquire), and to materials published in recent years. If your search yields too many citations, try specifying the search terms more precisely (e.g., “neighborhood social capital”). If you have not found much literature, try using more general or multiple terms (e.g., “social relations” or “social ties”). You may want to include the names of authors of major relevant studies or of the most important relevant journals in addition to the subject terms. Whatever terms you search first,

don’t consider your search complete until you have tried several different approaches and have seen how many articles you find. A search for “domestic violence” in SocINDEX on July 30, 2017, yielded 12,358 hits; by adding “effects” or “influences” as required search terms and limiting the search to peer-reviewed articles published since 2010, the number of hits dropped to 566. But focusing even more by adding “police response” resulted in just 2 articles. So if you are focusing on issues like those in the Sherman and Burk study, you probably need to use a strategy a bit narrower than the next-to-last one, but not as narrow as that last one.

5. *Use Boolean search logic.* It’s often a good idea to narrow your search by requiring that abstracts contain combinations of words or phrases that include more of the specific details of your research question. Using the Boolean connector *AND* allows you to do this, whereas using the connector *OR* allows you to find abstracts containing different words that mean the same thing (see Exhibit 2.5 later in this chapter).
6. *Use appropriate subject descriptors.* Once you have found an article that you consider appropriate, check the“descriptors” or “subject terms” field in the citation. You can then redo your search after requiring that the articles be classified with some or all of these descriptor terms. It can be helpful to add “key terms” that have already been attached to related articles by professional indexers (Gough et al. 2017:108–110). Such “controlled vocabularies” appear as “subject terms” in SocINDEX, as “thesaurus” terms in Sociological Abstracts, and as MeSH (medical subject headings) in MEDLINE. A librarian can help you to refine your search strategy in this way.
7. *Check the results.* Read the titles and abstracts you have found and identify the articles that appear to be most relevant. If possible, click on these article titles and generate a list of their references. See if you find more articles that are relevant to your research question but that you have missed so far. You will be surprised (I always am) at how many important articles your initial online search missed.
8. *Locate the articles.* Whatever database you use, the next step after finding your references is to obtain the articles themselves (or begin first by just selecting on the basis of the titles). You will probably find the full text of many articles available online, but this will be determined by what journals your library subscribes to and the period for which it pays for online access. The most recent issues of some journals may not be available online. Keep in mind that your library will not have anywhere near all the journals (and books) that you run across in your literature search, so you will have to add another step to your search: checking the “holdings” information.

   If an article that appears to be important for your topic isn’t available from your own library, or online, you may be able to request a copy online through your library site or by asking a member of the library staff. You can also check http://worldcat.org to see what other libraries have the journal.
9. *Take notes on each article you read, organizing your notes into standard sections: theory, methods, findings, conclusions.* In any case, write your review of the literature so that it contributes to your study in some concrete way; don’t feel compelled to discuss an

> article just because you have read it. Be judicious. You are conducting only one study of one issue, and it will only obscure the value of your study if you try to relate it to every tangential point in related research.

Don’t think of searching the literature as a one-time-only venture—something that you leave behind as you move on to your *real* research. You may encounter new questions or unanticipated problems as you conduct your research or as you burrow deeper into the literature. Searching the literature again to determine what others have found in response to these questions or what steps they have taken to resolve these problems can yield substantial improvements in your own research. There is so much literature on so many topics that it often is not possible to figure out in advance every subject for which you should search the literature, or what type of search will be most beneficial.

Another reason to make searching the literature an ongoing project is that the literature is always growing. During the course of one research study, whether it takes only one semester or several years, new findings will be published and relevant questions will be debated. Staying attuned to the literature and checking it at least when you are writing up your findings may save your study from being outdated as soon as it is finished.

Reviewing Research

Your literature review will suggest specific research questions for further investigation and research methods with which to study those questions. Sherman and Berk (1984) learned from their literature review that there had been little empirical research about the impact of arrest policies in domestic violence cases. What prior research had been conducted did not use very rigorous research designs. There was thus potential value in conducting new research using a rigorous design. Subsequent researchers questioned whether Sherman and Berk's results would be replicated in other locations and whether some of their methods could be improved. When the original results did not replicate, researchers designed more investigations to test explanations for the different findings. In this way, reviewing the literature identifies unanswered questions and contradictory evidence.

Effective review of the prior research is thus an essential step in building the foundation for new research. You must assess carefully the quality of each research study, consider the implications of each article for your own plans, and expand your thinking about your research question to account for new perspectives and alternative arguments. Through reviewing the literature and using it to extend and sharpen your own ideas and methods, you become a part of the social science community. Instead of being just one individual studying an issue that interests you, you are building on an ever-growing body of knowledge that is being constructed by the community of scholars. The value of the Sechrist and Weil (2017) study is lessened somewhat because they do not review the Sherman and Berk study or its replications, and thus do not compare their methods or findings to those of these other very relevant investigations.

Sometimes you'll find that someone else has already searched the literature on your research question and discussed what they found in a special review article or book chapter. For example, Aygül Akyüz, Tülay Yavan, Gönül Şahiner, and Ayşe Kılıç (2012) published a review of the research on domestic violence and women's reproductive health in the journal *Aggression and Violent Behavior.* Most of the research articles that you find will include a short literature review on the specific focus of the research. These reviews can help a lot, but they are no substitute for searching the literature yourself, selecting the articles and other sources that are most pertinent to your research question, and then reviewing what you have found. No one but you can decide what is relevant for your research question and the research circumstances you will be facing—the setting you will study, the timing of your study, the new issues that you want to include in your study, and your specific methods. And you can't depend on any published research review for information on the most recent works. New research results about many questions appear continually in scholarly journals and books, in research reports from government agencies and other organizations, and on websites all over the world; you'll need to check for new research like this yourself.

The published scholarly journal literature can be identified in databases such as Sociological Abstracts, SocINDEX, PsycINFO, and MEDLINE. Because these literature databases follow a standard format and use a careful process to decide what literature to include, they are the sources on which you should focus. This section concentrates on the procedures you should use for reviewing the articles you find in a search of the scholarly literature, but these procedures can also be applied to reviews of research monographs—books that provide more information from a research project than can be contained in a journal article.

Reviewing the literature is really a two-stage process. In the first stage, you must assess each article separately. This assessment should follow a standard format such as that represented by the "Questions to Ask About a Research Article" in Appendix A. However, keep in mind that you can't adequately understand a research study if you just treat it as a series of discrete steps involving a marriage of convenience among separate techniques. Any research project is an integrated whole, so you must be concerned with how each component of the research design influenced the others—for example, how the measurement approach might have affected the causal validity of the researcher's conclusions and how the sampling strategy might have altered the quality of measures.

The second stage of the review process is to assess the implications of the entire set of articles (and other materials) for the relevant aspects of your research question and procedures, and then to write an integrated narrative review that highlights these implications. Although you can find literature reviews that consist simply of assessments of one published article after another—that never get beyond the first stage in the review process—your understanding of the literature and the quality of your own work will be much improved if you make the effort to write an integrated review.

An alternative and increasingly popular approach to searching and reviewing the literature is to conduct a "systematic view." This approach begins with a very carefully designed search strategy and inclusion criteria, and then summarizes each included study on a detailed review protocol.

In the next section, I show how you might answer many of the questions in Appendix A as I review a research article about domestic violence. I will then show how the review of a single article can be used within an integrated review of the body of prior research on this research question. Because at this early point in the text you won't be familiar with all the terminology used in the article review, you might want to read through the more elaborate article review in Appendix B later in the course. I will conclude with a section on conducting a systematic review.

Single-Article Reviews: Formal and Informal Deterrents to Domestic Violence

Antony Pate and Edwin Hamilton at the national Police Foundation designed one of the studies funded by the U.S. Department of Justice to replicate the Minneapolis Domestic Violence Experiment. In this section, we will examine the article that resulted from that replication, which was published in the *American Sociological Review* (Pate and Hamilton 1992). The numbers in square brackets refer to the article review questions in Appendix A.

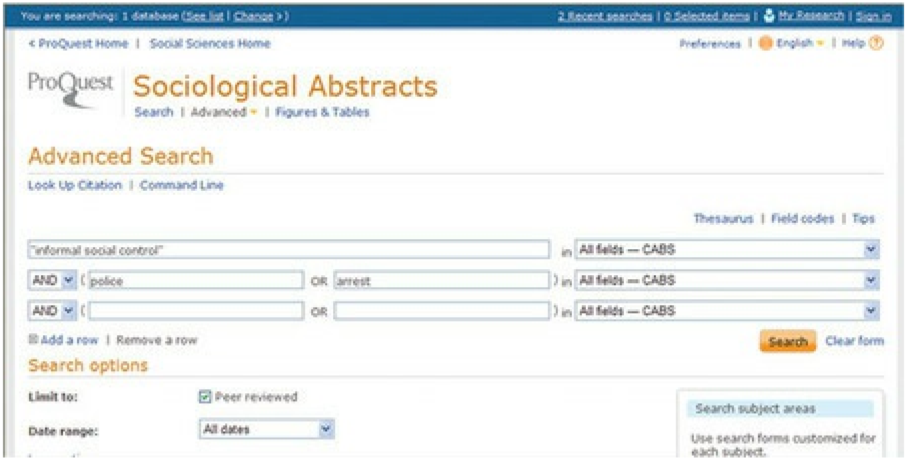
The research question.

Like Sherman and Berk's (1984) original Minneapolis study, Pate and Hamilton's (1992) Metro-Dade spouse assault experiment sought to test the deterrent effect of arrest in domestic violence cases, but with an additional focus on the role of informal social control [1]. The purpose of the study was explanatory, because the goal was to explain variation in the propensity to commit spouse abuse [2]. Deterrence theory provided the theoretical framework for the study, but this framework was broadened to include the proposition by Kirk Williams and Richard Hawkins (1986) that informal sanctions such as stigma and the loss of valued relationships augment the effect of formal sanctions such as arrest [4]. Pate and Hamilton's (1992) literature review referred, appropriately, to the original Sherman and Berk (1984) research, to the other studies that attempted to replicate the original findings, and to research on informal social control [3].

Exhibit 2.5 shows what Pate and Hamilton might have entered on their computer if they searched Sociological Abstracts to find research on "informal social control" and "police" or "arrest."

There is no explicit discussion of ethical guidelines in the article, although reference is made to a more complete unpublished report [6]. Clearly, important ethical issues had to be considered, given the experimental intervention in the police response to serious assaults, but the adherence to standard criminal justice procedures suggests attention to the welfare of victims as well as to the rights of suspects. We will consider issues in more detail in the next chapter.

Exhibit 2.5 Starting a Search in Sociological Abstracts


You are searching: 1 database (See list | Change >)
2 Recent searches | 0 Selected items | My Research | Sign in
< ProQuest Home | Social Sciences Home
Preferences | English | Help
ProQuest
Sociological Abstracts
Search | Advanced | Figures & Tables
Advanced Search
Look Up Citation | Command Line
Thesaurus | Field codes | Tips
Add a row | Remove a row
Search
Clear form
Search options
Search subject areas
Use search forms customized for each subject.

"informal social control" in All fields — CABS

AND (police OR arrest) in All fields — CABS

AND (OR) in All fields — CABS

Limit to: ☑ Peer reviewed

Date range: All dates

The research design.

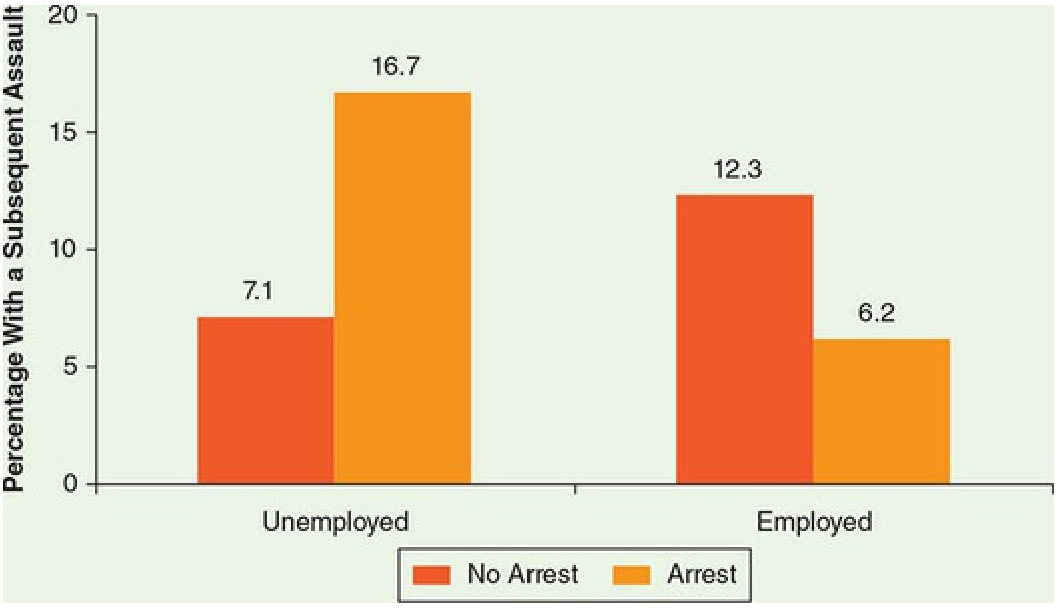
Developed as a follow-up to the original Minneapolis experiment, the Metro-Dade experiment exemplifies the features of a well-designed, deductive evaluation research study [5]. It was designed systematically, with careful attention to specification of terms and clarification of assumptions, and focused on the possibility of different outcomes rather than certainty about one preferred outcome. The major concepts in the study, formal and informal deterrence, were defined clearly [9] and then measured with straightforward indicators—arrest or nonarrest for formal deterrence and marital status and employment status for informal deterrence. However, the specific measurement procedures for marital and employment status were not discussed, and no attempt was made to determine whether they captured adequately the concept of informal social control [9, 10].

Three hypotheses were stated and related to the larger theoretical framework and prior research [7]. The study design focused on the behavior of individuals [13] and collected data over time, including records indicating subsequent assault as many as 6 months after the initial arrest [14]. The project's experimental design was used appropriately to test for the causal effect of arrest on recidivism [15, 17]. The research project involved all eligible cases, rather than a sample of cases, but there were a number of eligibility criteria that narrowed the ability to generalize these results to the entire population of domestic assault cases in the Metro-Dade area or elsewhere [11]. There is a brief discussion of the 92 eligible cases that were not given the treatment to which they were assigned, but it does not clarify the reasons for the misassignment [15].

The research findings and conclusion.

Pate and Hamilton's (1992) analysis of the Metro-Dade experiment was motivated by concern with the effect of social context because the replications of the original Minneapolis Domestic Violence Experiment in other locations had not had consistent results [19]. Pate and Hamilton's analysis gave strong support to the expectation that informal social control processes are important: As they had hypothesized, arrest had a deterrent effect on suspects who were employed, but not on those who were unemployed (see Exhibit 2.6). However, marital status had no such effect [20]. The subsequent discussion of these findings gives no attention to the implications of the lack of support for the effect of marital status [21], but the study represents an important improvement over earlier research that had not examined informal sanctions [22]. The need for additional research is highlighted, and the importance of the findings for social policy is discussed: Pate and Hamilton suggest that their finding that arrest deters only those who have something to lose (e.g., a job) must be considered when policing policies are established [23].

Exhibit 2.6 Percentage of Suspects With a Subsequent Assault by Employment Status and Arrest Status



Source: Pate, Antony M. and Hamilton Edwin E. 1992. “Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment.” American Sociological Review 57(October):691–697.

Overall, the Pate and Hamilton (1992) study represents an important contribution to understanding how informal social control processes influence the effectiveness of formal sanctions such as arrest. Although the use of a population of actual spouse assault cases precluded the use of very sophisticated measures of informal social control, the experimental design of the study and the researchers’ ability to interpret the results in the context of several other comparable experiments distinguish this research as exceptionally worthwhile. It is not hard to understand why these studies continue to stimulate further research and ongoing policy discussions.

Integrated Literature Reviews: When Does Arrest Matter?

The goal of the second stage of the literature review process is to integrate the results of your separate article reviews and develop an overall assessment of the implications of prior research. The integrated literature review should accomplish three goals: (1) summarize prior research, (2) critique prior research, and (3) present pertinent conclusions (Hart 1998:186–187). I’ll discuss each of these goals in turn.

1. *Summarize prior research.* Your summary of prior research must focus on the particular research questions that you will address, but you may also need to provide some more general background. Carolyn Hoyle and Andrew Sanders (2000:14) begin

their *British Journal of Criminology* research article about mandatory arrest policies in domestic violence cases with what they term a "provocative" question: What is the point of making it a crime for men to assault their female partners and ex-partners? The researchers then review the different theories and supporting research that have justified different police policies: the "victim choice" position, the "pro-arrest" position, and the "victim empowerment" position. Finally, Hoyle and Sanders review the research on the "controlling behaviors" of men that frames the specific research question on which they focus: how victims view the value of criminal justice interventions in their own cases (Hoyle and Sanders 2000:15).

Ask yourself three questions about your summary of the literature:

   a. Have you been selective? If there have been more than a few prior investigations of your research question, you will need to narrow your focus to the most relevant and highest-quality studies. Don't cite a large number of prior articles "just because they are there."
   b. Is the research up-to-date? Be sure to include the most recent research, not just the "classic" studies.
   c. Have you used direct quotes sparingly? To focus your literature review, you need to express the key points from prior research in your own words. Use direct quotes only when they are essential for making an important point (Pyrczak 2005:51–59).

2. *Critique prior research.* Evaluate the strengths and weaknesses of the prior research. In addition to all the points that you develop as you answer the article review questions in Appendix A, you should also select articles for review that reflect work published in peer-reviewed journals and written by credible authors who have been funded by reputable sources. Consider the following questions as you decide how much weight to give each article:
   a. How was the report reviewed before its publication or release? Articles published in academic journals go through a rigorous review process, usually involving careful criticism and revision. Top refereed journals may accept only 10% of the submitted articles, so they can be very selective. Dissertations go through a lengthy process of criticism and revision by a few members of the dissertation writer's home institution. A report released directly by a research organization is likely to have had only a limited review, although some research organizations maintain a rigorous internal review process (see discussion of research organizations, later in this chapter). Papers presented at professional meetings may have had little prior review. Needless to say, more confidence can be placed in research results that have been subject to a more rigorous review.
   b. What is the author's reputation? Reports by an author or team of authors who have published other work on the research question should have somewhat greater credibility at the outset.
   c. Who funded and sponsored the research? Major federal funding agencies and

private foundations fund only research proposals that have been evaluated carefully and ranked highly by a panel of experts. These agencies also often monitor closely the progress of the research. This does not guarantee that every such project report is good, but it goes a long way toward ensuring some worthwhile products. Conversely, research that is funded by organizations that have a preference for a particular outcome should be given particularly close scrutiny (Locke, Silverman, and Spirduso 1998:37–44).

3. *Present pertinent conclusions.* Don't leave the reader guessing about the implications of the prior research for your own investigation. Present the conclusions you draw from the research you have reviewed. As you do so, follow several simple guidelines:
    a. Distinguish clearly your own opinion of prior research from the conclusions of the authors of the articles you have reviewed.
    b. Make it clear when your own approach is based on the theoretical framework that you use and not on the results of prior research.
    c. Acknowledge the potential limitations of any empirical research project. Don't emphasize problems in prior research that you can't avoid (Pyrczak 2005:53–56).
    d. Explain how the unanswered questions raised by prior research or the limitations of methods used in prior research make it important for you to conduct your own investigation (Fink 2005:190–192).

A good example of how to conclude an integrated literature review is provided by an article based on the replication in Milwaukee of the Minneapolis Domestic Violence Experiment. For this article, the late Ray Paternoster and his colleagues (1997) sought to determine whether police officers' use of fair procedures when arresting assault suspects would lessen the rate of subsequent domestic violence. Paternoster et al. (1997) conclude that there has been a major gap in the prior literature: "Even at the end of some seven experiments and millions of dollars, then, there is a great deal of ambiguity surrounding the question of how arrest impacts future spouse assault" (p. 164). Specifically, the researchers note that each of the seven experiments focused on the effect of arrest itself, but ignored the possibility that "particular kinds of police procedure might inhibit the recurrence of spouse assault" (p. 165).

So, Paternoster and his colleagues (1997) ground their new analysis in additional literature on procedural justice and conclude that their new analysis will be "the first study to examine the effect of fairness judgments regarding a punitive criminal sanction (arrest) on serious criminal behavior (assaulting one's partner)" (p. 172).

Systematic Literature Reviews: Second Responder Programs and Repeat Family Abuse Incidents

Any literature review should be systematic, but the term systematic review designates an approach to literature review that is much more structured than what is usually done in a narrative review. It is best thought of as a distinct method of research, for it often is itself the basis of a published article—rather than being used to provide the background in an article that reports the results of a new research study. Professional librarians often play a major role in systematic reviews and may coauthor the resulting article. There are also helpful tools available at http://systematicreviewtools.com/index.php. and https://systematicreviewsjournal.biomedcentral.com/. Published systematic reviews are now archived on searchable websites, including http://www.cochranelibrary.com/ (health care), https://www.campbellcollaboration.org/library.html (social interventions), and http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=56 (multiple topic areas).

Systematic review: A literature review that "uses a specific and reproducible method to identify, select and appraise studies of a previously agreed level of quality (either including all studies or only those that pass a minimum quality threshold) that are relevant to a particular question." (Booth et al. 2016:11)

The systematic review approach can be defined formally as a literature review that "uses a specific and reproducible method to identify, select and appraise studies of a previously agreed level of quality (either including all studies or only those that pass a minimum quality threshold) that are relevant to a particular question" (Booth et al. 2016:11).

Developing a systematic review involves all of the activities I have just described for searching and reviewing the literature, but it proceeds with more explicit plans and reports at each stage of the process. Many systematic review efforts adhere to the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses): http://www.prisma-statement.org/Default.aspx. A systematic review of the effect of second responder programs (involving a social worker or specially trained officer who visits the victim after the initial police response) on repeat incidents of family abuse by Robert Davis, David Weisburd, and Bruce Taylor (2008) provides examples of such reports.

1. *Define one or more specific research questions.* One of the three posed by Davis, Weisburd, and Taylor was "Do second responder programs decrease or increase abuse as measured on victim surveys?"
2. *List the terms to be used in searching and the specific sources to be searched.* After a detailed description of the criteria for inclusion of studies, Davis et al. listed the following sources and terms to be searched:
    a. Criminal Justice Periodical Index
    b. Criminal Justice Abstracts
    c. National Criminal Justice Reference Services (NCJRS) Abstracts
    d. Sociological Abstracts
    e. Social Science Abstracts (SocialSciAbs)

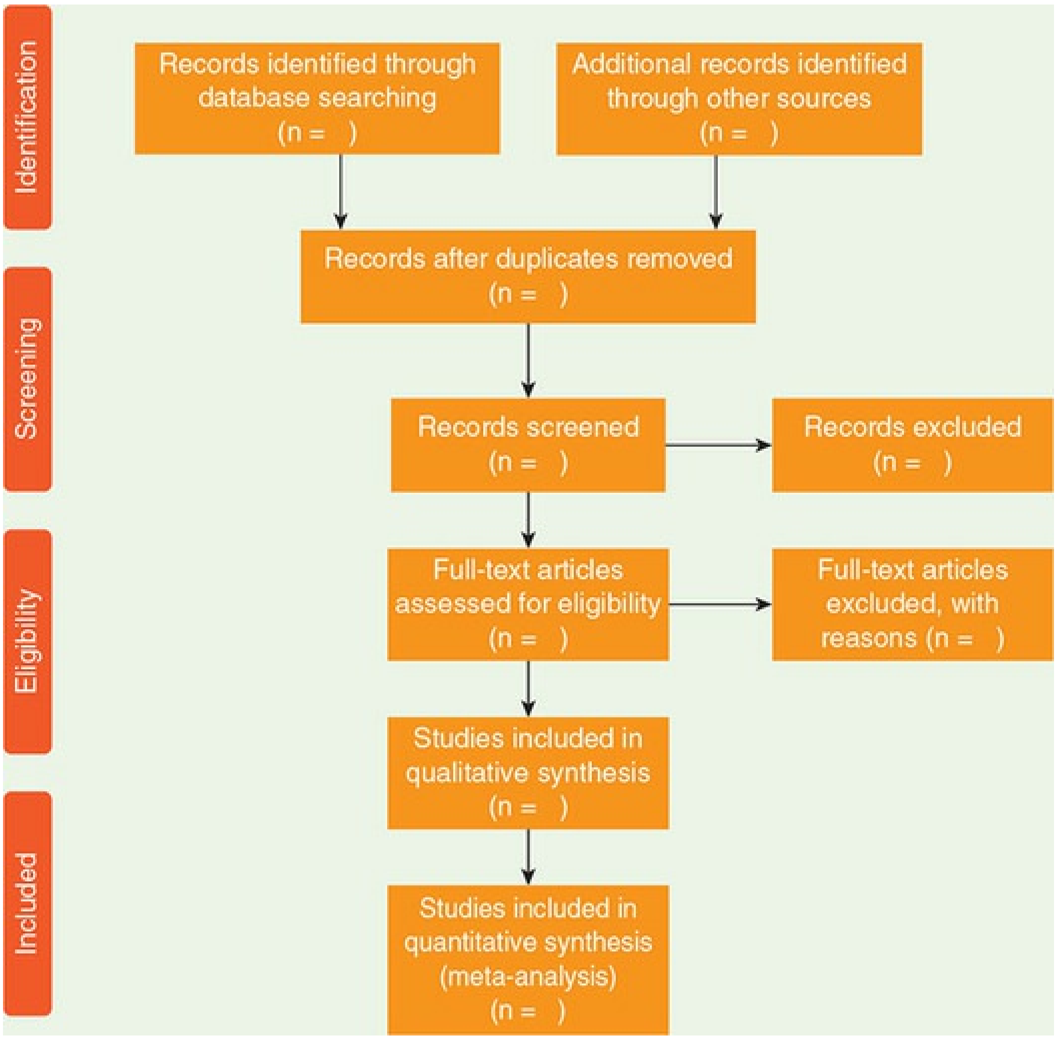
    f. Social Science Citation Index
    g. Dissertation Abstracts
    h. National Institute of Justice
    i. Office of Violence Against Women
    j. Office for Victims of Crime
    k. British Home Office
    l. Australian Criminology Database (CINCH)
    m. Government Publications Office, Monthly Catalog (GPO Monthly)
    n. C2 SPECTR (The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register)
    o. PsycINFO
    p. Google
    q. Google Scholar

   The following keywords were used to search the databases listed above (in all cases where police is listed, the word "policing" was also used):

    a. Second responder program
    b. Coordinated community response
    c. Police OR law enforcement AND repeat domestic violence OR wife abuse OR marital violence
    d. Police OR law enforcement AND crisis intervention AND domestic violence OR marital violence OR wife abuse
    e. Police OR law enforcement AND domestic violence advocates OR battered wom*n OR family violence AND evaluation AND response OR services
    f. Police OR law enforcement AND home visitation AND evaluation
    g. Police OR law enforcement AND intimate partner violence AND evaluation AND response OR services

3. *Report the results of the search and selection process,* often in the type of flow diagram recommended by PRISMA (see Exhibit 2.7).

Exhibit 2.7 PRISMA Flow Diagram


Identification
Screening
Eligibility
Included
Records identified through database searching (n =)
Additional records identified through other sources (n =)
Records after duplicates removed (n =)
Records screened (n =)
Records excluded (n =)
Full-text articles assessed for eligibility (n =)
Full-text articles excluded, with reasons (n =)
Studies included in qualitative synthesis (n =)
Studies included in quantitative synthesis (meta-analysis) (n =)

Source: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

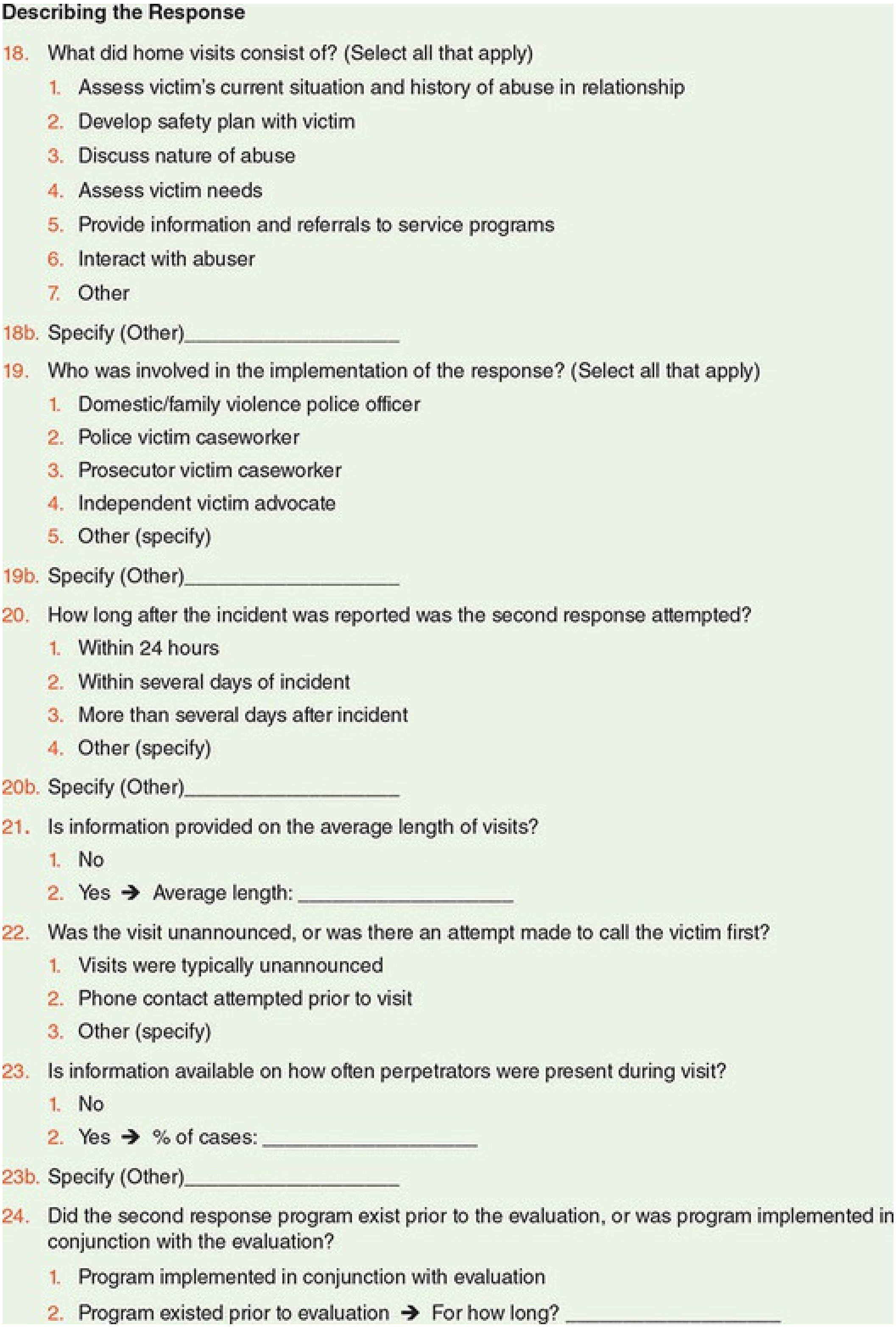
4. *Code the characteristics of the selected studies.* An excerpt from the coding sheets used by Davis et al. appears in Exhibit 2.8.
5. *Summarize the results in a narrative review.* A statistical evaluation of the outcomes of the studies is also included. The technique used to analyze statistically the composite results of multiple studies is termed *meta-analysis.* I will introduce this technique in Chapter 9. Davis, Weisburd, and Taylor (2008) found that for the 10 studies they reviewed, the second responder intervention increased slightly victims' willingness to report incidents to the police but did not affect the likelihood of new family violence incidents.

Systematic reviews provide an excellent source of information about prior research in areas in which interventions have been tested and such reviews have been conducted. Be sure to read any that are available pertaining to your research question.

Searching the Web

The World Wide Web provides access to vast amounts of information of many different sorts (Ó Dochartaigh 2002). There are more than 1 billion websites (http://www.internetlivestats.com/total-number-of-websites/) and by some calculations more than 4.6 billion webpages. (https://www.livescience.com/54094-how-big-is-the-internet.html). You can search the holdings of other libraries and download the complete text of government reports, some conference papers, and newspaper articles. You can find policies of local governments, descriptions of individual social scientists and particular research projects, and postings of advocacy groups. It's also hard to avoid finding a lot of information in which you have no interest, such as commercial advertisements, third-grade homework assignments, or advertisements galore.

Exhibit 2.8 Second Responder Meta-Analysis Coding Sheets: Excerpt (Davis et al. 2008)



Describing the Response

18. What did home visits consist of? (Select all that apply)
    1. Assess victim's current situation and history of abuse in relationship
    2. Develop safety plan with victim
    3. Discuss nature of abuse
    4. Assess victim needs
    5. Provide information and referrals to service programs
    6. Interact with abuser
    7. Other

18b. Specify (Other)____________________

19. Who was involved in the implementation of the response? (Select all that apply)
    1. Domestic/family violence police officer
    2. Police victim caseworker
    3. Prosecutor victim caseworker
    4. Independent victim advocate
    5. Other (specify)

19b. Specify (Other)____________________

20. How long after the incident was reported was the second response attempted?
    1. Within 24 hours
    2. Within several days of incident
    3. More than several days after incident
    4. Other (specify)

20b. Specify (Other)____________________

21. Is information provided on the average length of visits?
    1. No
    2. Yes → Average length: ____________________

22. Was the visit unannounced, or was there an attempt made to call the victim first?
    1. Visits were typically unannounced
    2. Phone contact attempted prior to visit
    3. Other (specify)

23. Is information available on how often perpetrators were present during visit?
    1. No
    2. Yes → % of cases: ____________________

23b. Specify (Other)____________________

24. Did the second response program exist prior to the evaluation, or was program implemented in conjunction with the evaluation?
    1. Program implemented in conjunction with evaluation
    2. Program existed prior to evaluation → For how long? ____________________

Source: Davis, R. C., Weisburd D, Taylor B. "Effects of second responder programs on repeat incidents of family abuse." Campbell Systematic Reviews 2008:15. © Davis et al.

So *caveat emptor* (buyer beware) is the watchword when you search the web. After all, it is a medium in which anyone with basic skills can post almost anything. On the other hand, the web provides access to resources of vital importance for many research projects; the trick is to use it well and wisely. Google has become the leading search engine. Its coverage is relatively comprehensive, and it does a good job of ranking search results by their relevancy (based on the terms in your search request). Google also allows you to focus your search just on images, discussions, or directories. There are several alternative search engines to Google, which each has some advantages and disadvantages (Gil 2017), but at this time most search needs can be met with Google.

Google Scholar is of special interest, since it provides a publicly accessible tool for searching the scholarly literature across disciplines—but also including technical reports, theses, books, and other types of documents. Google Scholar found 1,650 documents in a search for "police response to domestic violence" (on July 31, 2017), and since it lists articles in order of use of the search terms, frequency of citation, and other reasonable factors, the first several pages of citations provide a good way to identify potentially important omissions from your literature searches in bibliographic databases available at your library. However, in most cases you will still need to go through your library to obtain the full text of the articles that interest you (if your library subscribes to the source journals).

On the web, less is usually more. Limit your inspection of websites to the first few pages that turn up in your list (they're ranked by relevance). See what those first pages contain and then try to narrow your search by including some additional terms. Putting quotation marks around a phrase that you want to search will also help to limit your search—searching for "informal social control" on Google (on July 31, 2017) produced about 133,000 sites, compared with the roughly 3,860,000 sites retrieved when I omitted the quotation marks, so Google searched "informal" *and* "social" *and* "control." You can limit the search more by limiting it to sites with the desired phrase in the title. For instance, the search *allintitle:"Informal Social Control"* resulted in a dramatically smaller yield of results (1,670 in this case). You can also narrow your search by date or in other ways by using Google's settings = "Advanced Search" feature (147 webpages after limiting the search to the past 10 years in the English language and in the United States). If you are looking for graphical information such as a graph or a chart, you can limit your search to those pages that contain an image. On Google, this just requires clicking on the "Images" link located above the search box.

Exhibit 2.9 The Links Between Theory and Data

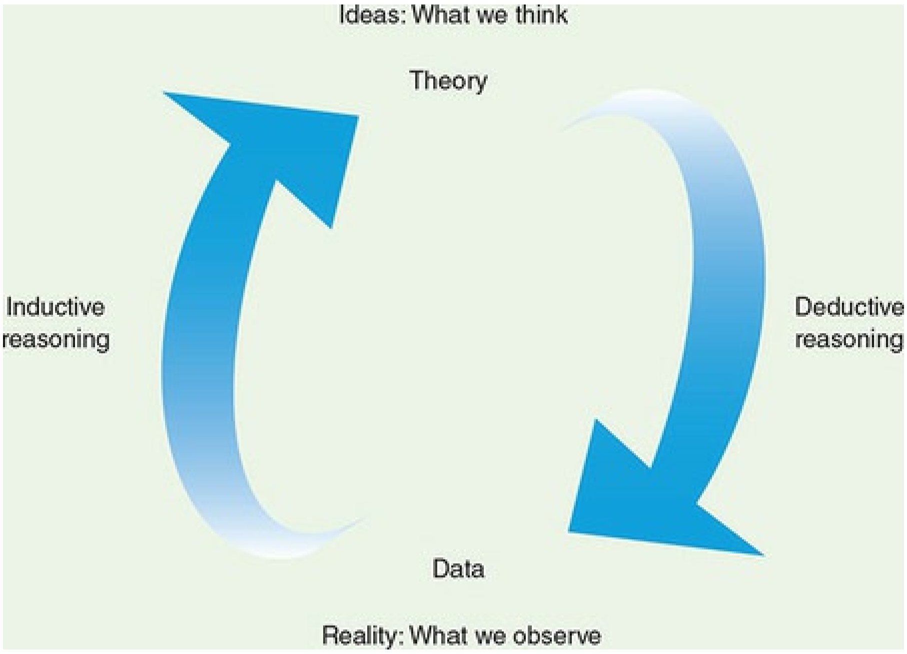
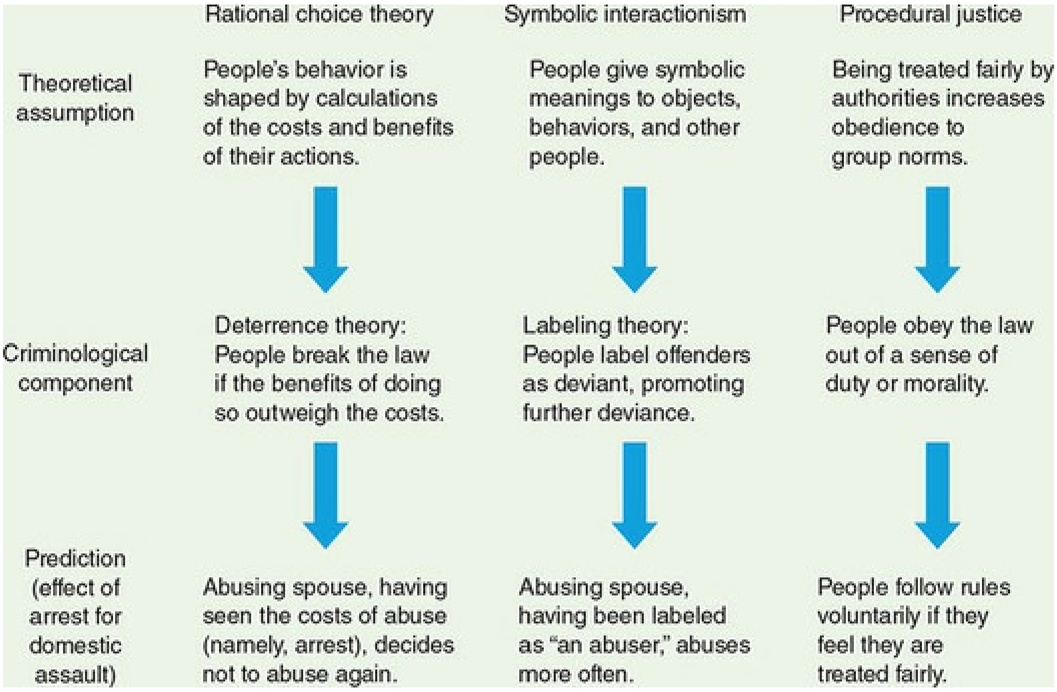


Exhibit 2.10 Three Social Theories and Their Predictions About the Effect of Arrest for Domestic Assault



	Rational choice theory	Symbolic interactionism	Procedural justice
Theoretical assumption	People's behavior is shaped by calculations of the costs and benefits of their actions.	People give symbolic meanings to objects, behaviors, and other people.	Being treated fairly by authorities increases obedience to group norms.
Criminological component	Deterrence theory: People break the law if the benefits of doing so outweigh the costs.	Labeling theory: People label offenders as deviant, promoting further deviance.	People obey the law out of a sense of duty or morality.
Prediction (effect of arrest for domestic assault)	Abusing spouse, having seen the costs of abuse (namely, arrest), decides not to abuse again.	Abusing spouse, having been labeled as "an abuser," abuses more often.	People follow rules voluntarily if they feel they are treated fairly.

Before you begin, be sure to clarify the goals of your search. Will you check on the coverage of your literature searching? Review related government programs? Find reports and statistics about the research question? Examine commentary about it? No matter what, be sure to record the URL (web address) for the useful sites you find.

Social Research Strategies

With a research question formulated, a review of the pertinent literature taking shape, and a theoretical framework in mind, you are ready to consider the process of conducting your research.

When you conduct social research, you are attempting to connect theory with empirical data—the evidence obtained from the social world. Researchers may make this connection by starting with a social theory and then testing some of its implications with data. This is the process of deductive research; it is most often the strategy used in quantitative methods. Alternatively, researchers may develop a connection between social theory and data by first collecting the data and then developing a theory that explains the patterns in the data (see Exhibit 2.9). This inductive research process is more often the strategy used in qualitative methods. As you'll see, a research project can draw on both deductive and inductive strategies.

In the News Research in the News: Control and Fear: What Mass Killings and Domestic Violence Have in Common

For Further Thought?

The June 2016 nightclub massacre in Orlando, Florida, was committed by a man, Omar Mateen, who had beaten his wife severely until she left him in 2009. FBI data reveal that, in 57% of mass shootings, a family member of the perpetrator is one of the victims, and social science research suggests that a desire for extreme control is a common factor in “intimate terrorism” and mass terrorism.

1. Does the proposal that these two forms of violence may stem from a similar underlying orientation make sense to you? Why or why not?
2. What type of research could improve our understanding of the possible link between domestic and mass violence?

News source: Taub, Amanda. 2016. “Control and Fear: What Mass Killings and Domestic Violence Have in Common.” *The New York Times,* June 15.

Social theories do not provide the answers to the questions we pose as topics for research. Instead, social theories suggest the areas on which we should focus and the propositions that we should consider for a test. Exhibit 2.10 summarizes how the two theories that guided Sherman and Berk’s (1984) research and the theory that guided Paternoster et al.’s (1997) reanalysis relate to the question of whether to arrest spouse abusers. By helping us make such connections, social theory sensitizes us to the possibilities, thus helping us design better research and draw out the implications of our results. Before, during, and after a research investigation, we need to keep thinking theoretically.

Explanatory Research

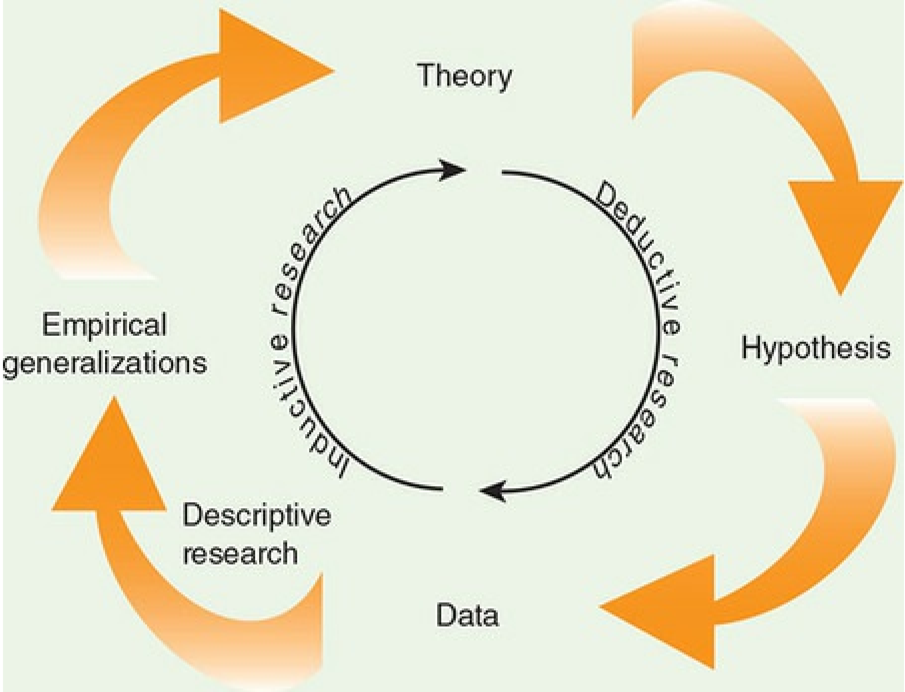
The process of conducting research designed to test explanations for social phenomena involves moving from theory to data and then back to theory. This process can be characterized with a research circle (Exhibit 2.11).

Research circle: A diagram of the elements of the research process, including theories, hypotheses, data collection, and data analysis.

Deductive Research

As Exhibit 2.11 shows, in deductive research, a specific expectation is deduced from a general theoretical premise and then tested with data that have been collected for this purpose. We call the specific expectation deduced from the more general theory a hypothesis. It is the hypothesis that researchers actually test, not the complete theory itself. A hypothesis proposes a relationship between two or more variables—characteristics or properties that can vary.

Exhibit 2.11 The Research Circle



Variation in one variable is proposed to predict, influence, or cause variation in the other. The proposed influence is the independent variable; its effect or consequence is the dependent variable. After the researchers formulate one or more hypotheses and develop research procedures, they collect data with which to test the hypothesis.

Hypotheses can be worded in several different ways, and identifying the independent and dependent variables is sometimes difficult. When in doubt, try to rephrase the hypothesis as an *if-then* statement: "*If* the independent variable increases (or decreases), *then* the dependent variable increases (or decreases)." Exhibit 2.12 presents several hypotheses with their independent and dependent variables and their if-then equivalents.

Exhibit 2.12 demonstrates another feature of hypotheses: direction of association. When researchers hypothesize that one variable increases as the other variable increases, the direction of association is positive (Hypotheses 1 and 4). When one variable decreases as the other variable decreases, the direction of association is also positive (Hypothesis 3). But when one variable increases as the other decreases, or vice versa, the direction of association is negative, or inverse (Hypothesis 2). Hypothesis 5 is a special case, in which the independent variable is qualitative: It cannot be said to increase or decrease. In this case, the concept of direction of association does not apply, and the hypothesis simply states that one category of the independent variable is associated with higher values on the dependent variable.

Both explanatory and evaluative studies are types of deductive research. The original Minneapolis Domestic Violence Experiment was an evaluative study because Sherman and Berk (1984) sought to explain what sort of response by the authorities might keep a spouse abuser from repeating the offense. The researchers deduced from deterrence theory the expectation that arrest would deter domestic violence. They then collected data to test this expectation.

In both explanatory and evaluative research, the statement of expectations for the findings and the design of the research to test these expectations strengthen the confidence we can place in the test. Deductive researchers show their hand or state their expectations in advance and then design a fair test of those expectations. Then, "the chips fall where they may"—in other words, the researcher accepts the resulting data as a more or less objective picture of reality.

Deductive research: The type of research in which a specific expectation is deduced from a general premise and is then tested.

Hypothesis: A tentative statement about empirical reality, involving a relationship between two or more variables.

Example of a hypothesis: The higher the poverty rate in a community, the higher the percentage of community residents who are homeless.

Variable: A characteristic or property that can vary (take on different values or attributes).

Example of a variable: The degree of honesty in verbal statements.

Independent variable: A variable that is hypothesized to cause, or lead to, variation in another variable.

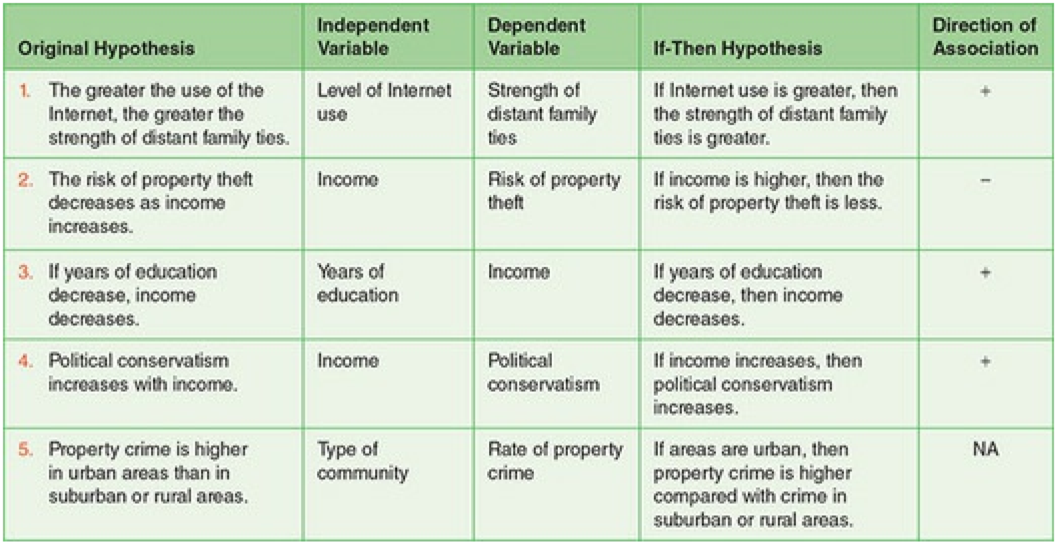
Example of an independent variable: Poverty rate.

Domestic Violence and the Research Circle

The classic Sherman and Berk (1984) study of domestic violence provides our first example of how the research circle works. In an attempt to determine ways to prevent the recurrence of spouse abuse, the researchers repeatedly linked theory and data, developing both hypotheses and empirical generalizations.

The first phase of Sherman and Berk's study was designed to test a hypothesis. According to deterrence theory, punishment will reduce recidivism, or the propensity to commit further crimes. From this theory, Sherman and Berk deduced a specific hypothesis: "Arrest for spouse abuse reduces the risk of repeat offenses." In this hypothesis, arrest is the independent variable and the risk of repeat offenses is the dependent variable (it is hypothesized to depend on arrest).

Exhibit 2.12 Examples of Hypotheses



Original Hypothesis	Independent Variable	Dependent Variable	If-Then Hypothesis	Direction of Association
1. The greater the use of the Internet, the greater the strength of distant family ties.	Level of Internet use	Strength of distant family ties	If Internet use is greater, then the strength of distant family ties is greater.	+
2. The risk of property theft decreases as income increases.	Income	Risk of property theft	If income is higher, then the risk of property theft is less.	–
3. If years of education decrease, income decreases.	Years of education	Income	If years of education decrease, then income decreases.	+
4. Political conservatism increases with income.	Income	Political conservatism	If income increases, then political conservatism increases.	+
5. Property crime is higher in urban areas than in suburban or rural areas.	Type of community	Rate of property crime	If areas are urban, then property crime is higher compared with crime in suburban or rural areas.	NA

Of course, in another study, arrest might be the dependent variable in relation to some other independent variable. For example, in the hypothesis "The greater the rate of layoffs in a community, the higher the frequency of arrest," the dependent variable is frequency of arrest. Only within the context of a hypothesis, or a relationship between variables, does it make sense to refer to one variable as dependent and the other as independent.

Sherman and Berk tested their hypothesis by setting up an experiment in which the police responded to the complaints of spouse abuse in one of three ways: (1) arresting the offender, (2) separating the spouses without making an arrest, or (3) simply warning the offender. When the researchers examined their data (police records for the persons in their experiment), they found that of those arrested for assaulting their spouse, only 13% repeated the offense, compared with a 26% recidivism rate for those who were separated from their spouse by the police without any arrest. This pattern in the data, or empirical generalization, was consistent with the hypothesis that the researchers deduced from deterrence theory. The theory thus received support from the experiment (see Exhibit 2.13).

Dependent variable: A variable that is hypothesized to vary depending on, or under the influence of, another variable.

Example of a dependent variable: Percentage of community residents who are homeless.

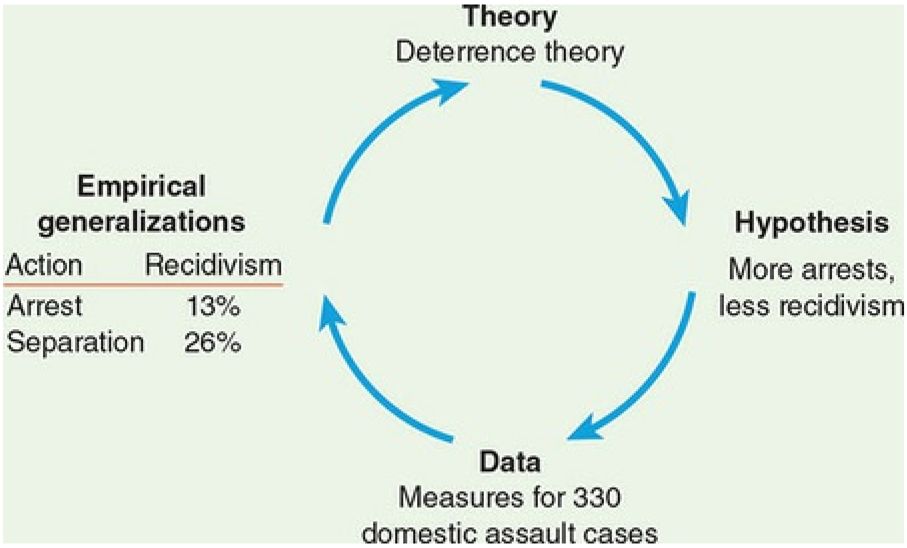
Direction of association: A pattern in a relationship between two variables—the values of variables tend to change consistently in relation to change on the other variable; the direction of association can be either positive or negative.

Empirical generalization: A statement that describes patterns found in data.

Because of their doubts about the generalizability of their results, Sherman, Berk, and other researchers began to journey around the research circle again, with funding from the National Institute of Justice for replications (repetitions) of the experiment in six more locations. These replications used the same basic research approach but with some improvements. The random assignment process was tightened in most of the locations so that police officers would be less likely to replace the assigned treatment with a treatment of their own choice. In addition, data were collected about repeat violence against other victims as well as against the original complainant. Some of the replications also examined different aspects of the arrest process, to see whether professional counseling helped and whether the length of time spent in jail after the arrest mattered at all.

Replications: Repetitions of a study using the same research methods to answer the same research question.

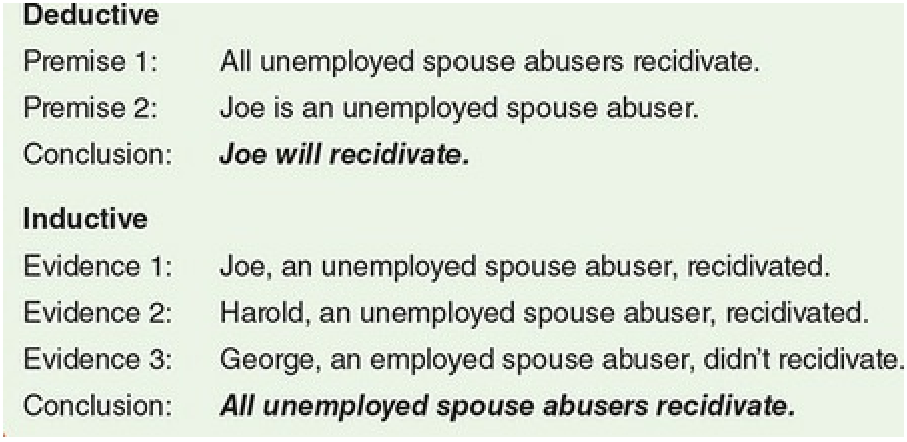
Exhibit 2.13 The Research Circle: Minneapolis Domestic Violence Experiment


Theory
Deterrence theory
Hypothesis
More arrests,
less recidivism
Data
Measures for 330
domestic assault cases
Empirical
generalizations
Action Recidivism
Arrest 13%
Separation 26%

Source: Data from Sherman and Berk 1984:267.

By the time results were reported from five of the locations in the new studies, a problem was apparent. In three of the locations—Omaha, Nebraska; Charlotte, North Carolina; and Milwaukee, Wisconsin—researchers were finding long-term increases in domestic violence incidents among arrestees. But in two—Colorado Springs, Colorado; and Dade County, Florida—the predicted deterrent effects seemed to be occurring (Sherman and Smith 1992). Sherman and his colleagues had now traversed the research circle twice in an attempt to answer the original research question, first in Minneapolis and then in six other locations. But rather than leading to more confidence in deterrence theory, the research results were questioning it. Deterrence theory now seemed inadequate to explain empirical reality, at least as the researchers had measured this reality. So, the researchers began to reanalyze the follow-up data from several locations in an attempt to explain the discrepant results, thereby starting around the research circle once again (Berk et al. 1992; Pate and Hamilton 1992; Sherman and Smith 1992).

Exhibit 2.14 Deductive and Inductive Reasoning



Deductive

Premise 1:	All unemployed spouse abusers recidivate.
Premise 2:	Joe is an unemployed spouse abuser.
Conclusion:	***Joe will recidivate.***

Inductive

Evidence 1:	Joe, an unemployed spouse abuser, recidivated.
Evidence 2:	Harold, an unemployed spouse abuser, recidivated.
Evidence 3:	George, an employed spouse abuser, didn't recidivate.
Conclusion:	***All unemployed spouse abusers recidivate.***

Inductive Research

In contrast to deductive research, inductive research begins with specific data, which are then used to develop (induce) a general explanation (a theory) to account for the data. One way to think of this process is in terms of the research circle: Rather than starting at the top of the circle with a theory, the inductive researcher starts at the bottom of the circle with data and then develops the theory. Another way to think of this process is represented in Exhibit 2.14. In deductive research, reasoning from specific premises results in a conclusion that a theory is supported, but in inductive research, the identification of similar empirical patterns results in a generalization about some social process.

Inductive reasoning enters into deductive research when we find unexpected patterns in the data we have collected for testing a hypothesis. We may call these patterns anomalous findings. When these unexpected patterns lead to new explanations, insights, or theoretical approaches, we call them serendipitous findings. However, the adequacy of an explanation formulated after the fact is necessarily less certain than an explanation presented before the collection of data and tested in a planned way with the data. Every phenomenon can always be explained in *some* way. Inductive explanations are thus more trustworthy if they are confirmed subsequently with deductive research.

Inductive research: The type of research in which general conclusions are drawn from specific data.

Anomalous findings: Unexpected patterns in data.

Serendipitous findings: Unexpected patterns in data, which stimulate new explanations, insights, or theoretical approaches.

An inductive approach to explaining domestic violence.

The domestic violence research took an inductive turn when Sherman and the other researchers began trying to make sense of the differing patterns in the data collected in the different locations. Could systematic differences in the samples or in the implementation of arrest policies explain the differing outcomes? Or was the problem an inadequacy in the theoretical basis of their research? Was deterrence theory really the best way to explain the patterns in the data they were collecting?

As you learned in my review of the Pate and Hamilton (1992) study, the researchers had found that individuals who were married and employed were deterred from repeat offenses by arrest, but individuals who were unmarried and unemployed were actually more likely to commit repeat offenses if they were arrested. What could explain this empirical pattern? The researchers turned to control theory, which predicts that having a "stake in conformity" (resulting from inclusion in social networks at work or in the community) decreases a person’s likelihood of committing crimes (Toby 1957). The implication is that people who are employed and married are more likely to be deterred by the threat of arrest than are those without such stakes in conformity. And this is indeed what the data revealed.

Now, the researchers had traversed the research circle almost three times, a process perhaps better described as a spiral (see Exhibit 2.15). The first two times, the researchers had traversed the research circle in a deductive, hypothesis-testing way. They started with theory and then deduced and tested hypotheses. The third time, they were more inductive: They started with empirical generalizations from the data they had already obtained and then turned to a new theory to account for the unexpected patterns in the data. At this point, they believed that deterrence theory made correct predictions, given certain conditions, and that another theory, control theory, might specify what these conditions were.

This last inductive step in their research made for a more complex, but also conceptually richer, picture of the impact of arrest on domestic violence. The researchers seemed to have come closer to understanding how to inhibit domestic violence. But they cautioned us that their initial question—the research problem—was still not completely answered. Employment status and marital status do not solely measure the strength of social attachments; they are also related to how much people earn and the social standing of victims in court. So, maybe social ties are not really what make arrest an effective deterrent to domestic violence. The real deterrent may be cost–benefit calculations (“If I have a higher income, jail is more costly for me”) or perceptions about the actions of authorities (“If I am a married woman, judges will treat my complaint more seriously”). Additional research was needed (Berk et al. 1992).

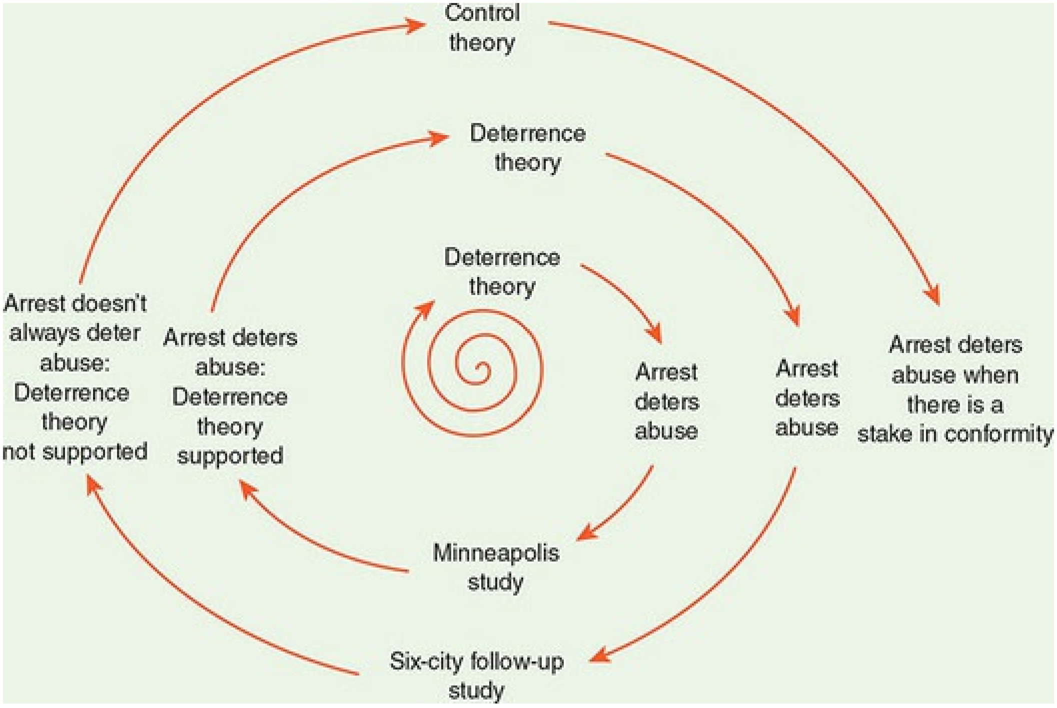
Exploratory Research

Qualitative research is often exploratory and, hence, inductive: The researchers begin by observing social interaction or interviewing social actors in depth and then developing an explanation for what has been found. The researchers often ask questions such as “What is going on here?” “How do people interpret these experiences?” or “Why do people do what they do?” Rather than testing a hypothesis, the researchers are trying to make sense of some social phenomenon. They may even put off formulating a research question until after they begin to collect data—the idea is to let the question emerge from the situation itself (Brewer and Hunter 1989:54–58).

Battered Women’s Help Seeking

Angela Moe (2007) used exploratory research methods in her study of women’s decisions to seek help after abuse experiences. Rather than interviewing women in court, Moe interviewed 19 women in a domestic violence shelter. In interviews lasting about 1 hour each, the women were able to discuss, in their own words, what they had experienced and how they had responded. Moe then reviewed the interview transcripts carefully and identified major themes that emerged in the comments.

Exhibit 2.15 The Research Spiral: Domestic Violence Experiment


Control theory
Deterrence theory
Deterrence theory
Arrest doesn't always deter abuse: Deterrence theory not supported
Arrest deters abuse: Deterrence theory supported
Arrest deters abuse
Arrest deters abuse
Arrest deters abuse when there is a stake in conformity
Minneapolis study
Six-city follow-up study

The following quote is from a woman who had decided not to call the police to report her experience of abuse (Moe 2007:686). We can use this type of information to identify some of the factors behind the underreporting of domestic violence incidents. Moe or other researchers might then design a survey of a larger sample to determine how frequently each basis for underreporting occurs.

> I tried the last time to call the police and he ripped both the phones out of the walls. . . .
>
> That time he sat on my upper body and had his thumbs in my eyes and he was just squeezing.
>
> He was going, "I'll gouge your eyes out. I'll break every bone in your body. Even if they do find you alive, you won't know to tell them who did it to you because you'll be in intensive care for so long you'll forget." (Terri)

The Moe (2007) example illustrates how the research questions that serve as starting points for qualitative data analyses do not simply emerge from the setting studied, but are shaped by the investigator. The research question can change, narrow, expand, or multiply throughout the processes of data collection and analysis.

Explanations developed inductively from qualitative research can feel authentic because we have heard what people have to say in their own words, and we have tried to see the social world as they see it. Explanations derived from qualitative research will be richer and more finely textured than they often are in quantitative research, but they are likely to be based on fewer cases from a limited area. We cannot assume that the people studied in this setting are like others or that other researchers will develop explanations similar to ours to make sense of what was observed or heard. Because we do not initially set up a test of a hypothesis according to some specific rules, another researcher cannot come along and conduct the same test.

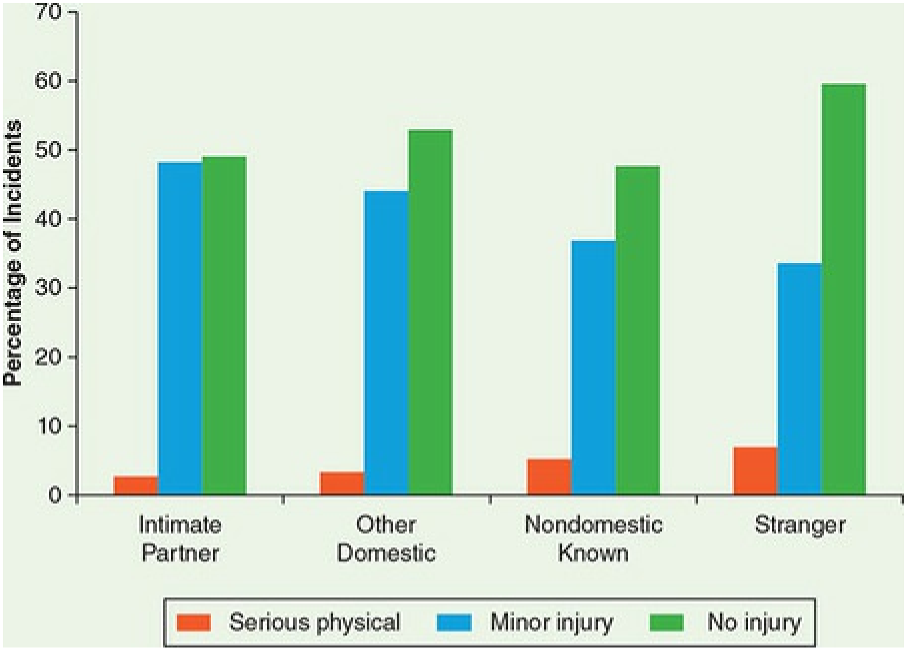
Descriptive Research

You learned in Chapter 1 that some social research is purely descriptive. Such research does not involve connecting theory and data, but it is still a part of the research circle—it begins with data and proceeds only to the stage of making empirical generalizations based on those data (refer to Exhibit 2.11).

Valid description is important in its own right—it is a necessary component of all investigations. Before they began an investigation of differences in arrests for domestic violence in states with and without mandatory arrest laws, David Hirschel, Eve Buzawa, April Pattavina, and Don Faggiani (2008) carefully described the characteristics of incidents reported to the police (see Exhibit 2.16). Describing the prevalence of intimate partner violence is an important first step for societies that seek to respond to this problem (refer to Exhibit 2.1). Government agencies and nonprofit organizations frequently sponsor research that is primarily descriptive: How many poor people live in this community? Is the health of the elderly improving? How frequently do convicted criminals return to crime? Simply put, good description of data is the cornerstone of the scientific research process and an essential component for understanding the social world.

Good descriptive research can also stimulate more ambitious deductive and inductive research. The Minneapolis Domestic Violence Experiment was motivated, in part, by a growing body of descriptive research indicating that spouse abuse is very common: 572,000 cases of women victimized by a violent partner each year (Buzawa and Buzawa 1996:1).

Exhibit 2.16 Incident, Offender, and Outcome Variables by Victim–Offender Relationship



Source: Based on Hirschel, David, Eve Buzawa, April Pattavina, and Don Faggiani. 2008. “Domestic Violence and Mandatory Arrest Laws: To What Extent Do They Influence Police Arrest Decisions?” Journal of Criminal Law & Criminology 98(1):255–298.

Careers and Research

Kristin M. Curtis, Senior Research Program Coordinator

Kristin Curtis graduated with a master’s degree in criminal justice from Rutgers University in Camden in 2010. As a graduate student, she worked on a nationwide research project examining policy maker and practitioner perspectives on sex offender laws, and this experience convinced her that pursuing a career in research was the best fit for her interests and talents. She secured a position as a graduate project assistant at a research institute where she worked on statewide prisoner reentry studies. Curtis has quickly moved up the ranks and in the process has worked on myriad criminal justice projects. Her research assignments require varied methodological approaches including interviews, focus groups, surveys, network analysis, regression models, and geographic information systems (GIS).

One feature of her work that Curtis truly values is the fact that she can participate in other areas of study outside the criminal justice realm. For instance, she has worked on projects that examine the impact of social service organization collaboration on child well-being, financial stability of families, and relationships between children and their caregivers. These projects involve the evaluation of collaborations among social service organizations in multiple counties and employ both qualitative and quantitative research methods. After 8 years, Curtis still enjoys her position as each day presents new challenges and different tasks, including collecting and analyzing data, finalizing reports, writing grant proposals for potential new projects, and supervising graduate students.

Curtis has advice for students interested in careers conducting research or using research results:

> Locate faculty who engage in research in your areas of interest. Even if you are unsure what your primary research areas are, working on a research project allows you to gain exposure to different research methodologies and techniques (i.e., quantitative and qualitative). You might find you enjoy research and pick up conference presentations and academic publications along the way. Remember, college is an opportunity to explore the different career choices in the world, so take

advantage of this.

Social Research Organizations

Another important consideration when evaluating or launching a research project is to consider how it is organized (Fawcett and Pockett 2015:14–16, 145–146). Will you be working with a mentor (your professor, perhaps) who will help you to make decisions and guide you toward the relevant literature? Will you be working as part of a team of equals with whom you can bounce around ideas and share tasks? Will others be hired to conduct interviews, enter data, consult on statistics? Did the research project involve a partnership between researchers and practitioners—who might have helped to shape interpretations of the results? Were there multiple research sites, where procedures might have been implemented somewhat differently? Is the project embedded within a larger service organization, such as an academic hospital or an introductory university course, that might have influenced the participants? Understanding the organizational context in which the research was or will be conducted will sensitize you to potential problems and alternative approaches.

Social Research Standards

Social science research can improve our understanding of empirical reality—the reality we encounter firsthand. We have achieved the goal of validity when our conclusions about this empirical reality are correct. I look out my window and observe that it is raining—a valid observation, if my eyes and ears are to be trusted. I pick up the newspaper and read that the rate of violence may be climbing after several years of decline. I am less certain of the validity of this statement, based as it is on an interpretation of some trends in crime indicators obtained through some process that isn't explained. As you learned in this chapter, many social scientists who have studied the police response to domestic violence came to the conclusion that arrest deters violence—that there is a valid connection between this prediction of rational choice theory and the data obtained in research about these processes.

Validity: The state that exists when statements or conclusions about empirical reality are correct.

If validity sounds desirable to you, you're a good candidate for becoming a social scientist. If you recognize that validity is often a difficult goal to achieve, you may be tough enough for social research. In any case, the goal of social science is not to come up with conclusions that people will like or conclusions that suit our own personal preferences. The goal is to figure out how and why the social world—some aspect of it—operates as it does. In *Investigating the Social World,* we are concerned with three standards for validity: (1) measurement validity, (2) generalizability, and (3) causal validity (also known as internal validity) (Hammersley 2008:43). We will learn that invalid measures, invalid generalizations, or invalid causal inferences will result in invalid conclusions. We will also focus on the standard of authenticity, a concern with reflecting fairly the perspectives of participants in a setting that we study.

Measurement validity: Exists when a measure measures what we think it measures.

Generalizability: Exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify.

Causal validity (internal validity): Exists when a conclusion that A leads to or results in B is correct.

Authenticity: When the understanding of a social process or social setting is one that reflects fairly the various perspectives of participants in that setting.

Measurement Validity

Measurement validity is our first concern in establishing the validity of research results because without having measured what we think we measured, we really don't know what we're talking about. Measurement validity is the focus of Chapter 4. A measure is valid when it measures what we think it measures. In other words, if we seek to describe the frequency of domestic violence in families, we need to develop a valid procedure for measuring domestic violence.

The first step in achieving measurement validity is to specify clearly what it is we intend to measure. Patricia Tjaden and Nancy Thoennes (2000) identified this as one of the problems with research on domestic violence: "definitions of the term vary widely from study to study, making comparisons difficult" (p. 5). To avoid this problem, Tjaden and Thoennes (2000) presented a clear definition of what they meant by *intimate partner violence:*

> Rape, physical assault, and stalking perpetrated by current and former dates, spouses, and cohabiting partners, with cohabiting meaning living together at least some of the time as a couple. (p. 5)

Tjaden and Thoennes also provided a measure of each type of violence. For example, "'physical assault' is defined as behaviors that threaten, attempt, or actually inflict physical harm" (Tjaden and Thoennes 2000:5).

With this definition in mind, Tjaden and Thoennes (2000:6) then specified the set of questions they would use to measure intimate partner violence (the questions pertaining to physical assault):

> Not counting any incidents you have already mentioned, after you became an adult, did any other adult, male or female, ever:
>
> - Throw something at you that could hurt?
> - Push, grab, or shove you?
> - Pull your hair?
> - Slap or hit you?
> - Kick or bite you?
> - Choke or attempt to drown you?
> - Hit you with some object?
> - Beat you up?
> - Threaten you with a gun?

- Threaten you with a knife or other weapon?
- Use a gun on you?
- Use a knife or other weapon on you?

Do you believe that answers to these questions provide a valid measure of having been physically assaulted? Do you worry that some survey respondents might not report all the assaults they have experienced? Might some respondents make up some incidents? Issues like these must be considered when we evaluate measurement validity. Suffice it to say that we must be very careful in designing our measures and in subsequently evaluating how well they have performed. Chapter 4 introduces several different ways to test measurement validity. We cannot just *assume* that measures are valid.

Generalizability

The generalizability of a study is the extent to which it can be used to inform us about persons, places, or events that were not studied. Generalizability is the focus of Chapter 5.

You have already learned in this chapter that Sherman and Berk's findings in Minneapolis about the police response to domestic violence simply did not hold up in several other locations: The initial results could not be generalized. As you know, this led to additional research to figure out what accounted for the different patterns in different locations.

If every person or community we study were like every other one, generalizations based on observations of a small number would be valid. But that's not the case. We are on solid ground if we question the generalizability of statements about research based on the results of a restricted sample of the population or in just one community or other social context.

Generalizability has two aspects. Sample generalizability refers to the ability to generalize from a sample, or subset, of a larger population to that population itself. This is the most common meaning of generalizability. Cross-population generalizability refers to the ability to generalize from findings about one group, population, or setting to other groups, populations, or settings (see Exhibit 2.17). Cross-population generalizability can also be referred to as external validity. (Some social scientists equate the term *external validity* to *generalizability,* but in this book, I restrict its use to the more limited notion of cross-population generalizability.)

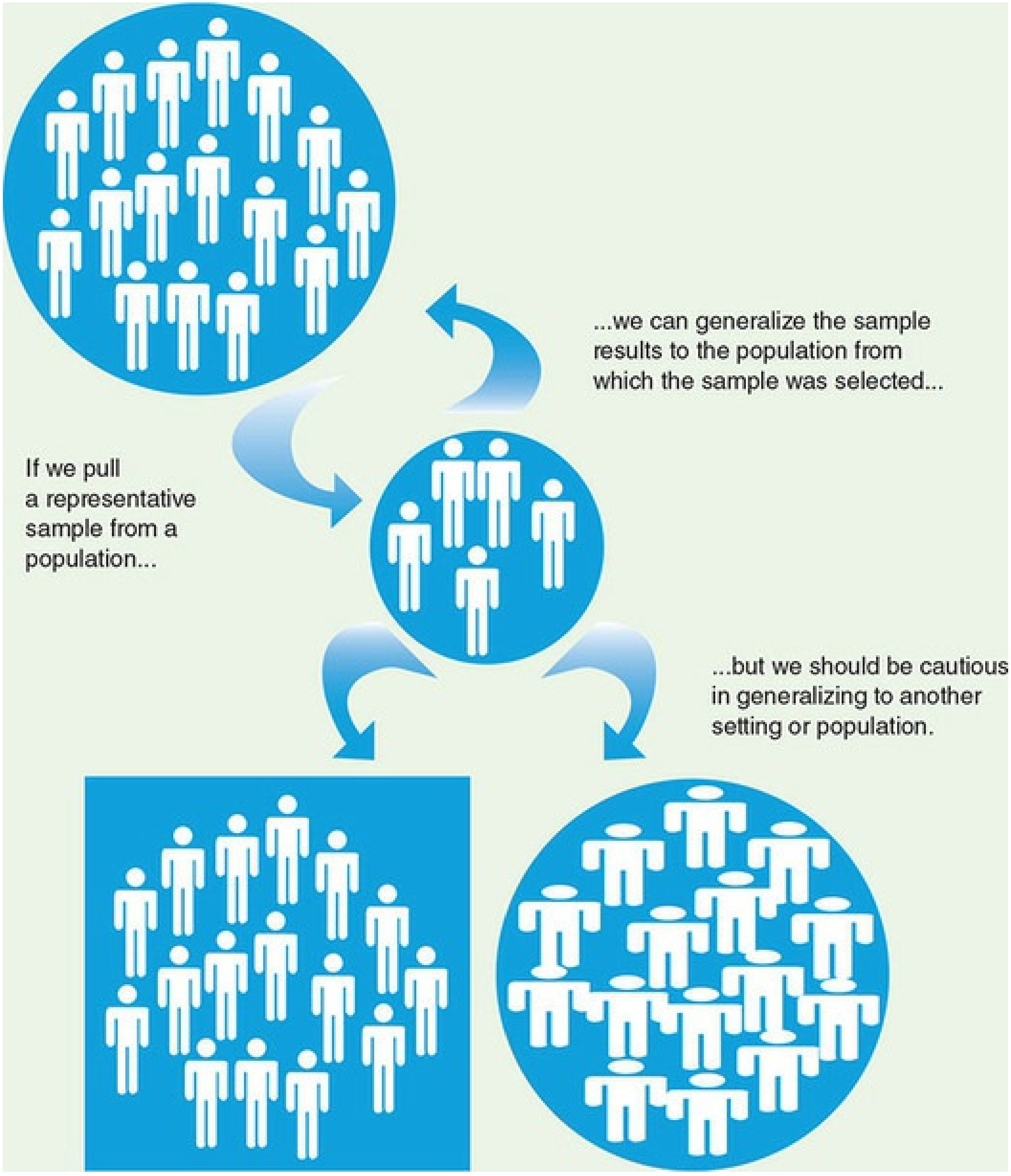
Sample generalizability: Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

Cross-population generalizability (external validity): Exists when findings about one group, population, or setting hold true for other groups, populations, or settings.

Sample generalizability is a key concern in survey research. Political pollsters may study a sample of likely voters, for example, and then generalize their findings to the entire population of likely voters. No one would be interested in the results of political polls if they represented only the relatively tiny sample that actually was surveyed rather than the entire population. The procedures for the National Violence Against Women Survey that Tjaden and Thoennes (2000) relied on were designed to maximize sample generalizability.

Cross-population generalizability occurs to the extent that the results of a study hold true for multiple populations; these populations may not all have been sampled, or they may be represented as subgroups within the sample studied. This was the problem with Sherman and Berk's (1984) results: Persons in Minneapolis who were arrested for domestic violence did not respond in the same way as persons arrested for the same crime in several other locations. The conclusions from Sherman and Berk's (1984) initial research in Minneapolis were not "externally valid."

Exhibit 2.17 Sample and Cross-Population Generalizability


If we pull a representative sample from a population...
...we can generalize the sample results to the population from which the sample was selected...
...but we should be cautious in generalizing to another setting or population.

Generalizability is a key concern in research design. We rarely have the resources to study the entire population that is of interest to us, so we have to select cases to study that will allow our findings to be generalized to the population of interest. Chapter 5 reviews alternative approaches to selecting cases so that findings can be generalized to the population from which the cases were selected. Nonetheless, because we can never be sure that our findings will hold under all conditions, we should be cautious in generalizing to populations or periods that we did not sample.

Causal Validity

Causal validity, also known as *internal validity,* refers to the truthfulness of an assertion that A causes B. It is the focus of Chapter 6.

Most research seeks to determine what causes what, so social scientists frequently must be concerned with causal validity. Sherman and Berk (1984) were concerned with the effect of arrest on the likelihood of recidivism by people accused of domestic violence. To test their causal hypothesis, Sherman and Berk designed their experiment so that some accused persons were arrested and others were not. Of course, it may seem heavy-handed for social scientists to influence police actions for the purpose of a research project, but this step reflects just how difficult it can be to establish causally valid understandings about the social world. Only because police officials did not know whether arrest caused spouse abusers to reduce their level of abuse were they willing to allow an experiment to test the effect of different policies. Hirschel and his collaborators (2008) used a different approach to investigate the effect of mandatory arrest laws on police decisions to arrest: They compared the rate of arrest for domestic violence incidents in jurisdictions with and without mandatory arrest laws.

Which of these two research designs gives you more confidence in the causal validity of the conclusions? Chapter 6 will give you much more understanding of how some features of a research design can help us evaluate causal propositions. However, you will also learn that the solutions are neither easy nor perfect: We always have to consider critically the validity of causal statements that we hear or read.

Authenticity

The goal of authenticity is stressed by researchers who focus attention on the subjective dimension of the social world. An authentic understanding of a social process or social setting is one that reflects fairly the various perspectives of participants in that setting (Gubrium and Holstein 1997). Authenticity is one of several different standards proposed by some observers as uniquely suited to qualitative research; it reflects a belief that those who study the social world should focus first and foremost on how participants view that social world, rather than on developing a unique social scientists' interpretation of that world. Rather than expecting social scientists to be able to provide a valid mirror of reality, this perspective emphasizes the need for recognizing that what is understood by participants as reality is a linguistic and social construction of reality (Kvale 2002:306).

Moe (2007) explained her basis for considering the responses of women she interviewed in the domestic violence shelter to be authentic:

> Members of marginalized groups are better positioned than members of socially dominant groups to describe the ways in which the world is organized according to the oppressions they experience. (p. 682)

Moe's (2007) assumption was that "battered women serve as experts of their own lives" (p. 682). Adding to her assessment of authenticity, Moe (2007) found that the women "exhibited a great deal of comfort through their honesty and candor" as they produced "a richly detailed and descriptive set of narratives" (p. 683). You will learn more about how authenticity can be achieved in qualitative methods in Chapters 10 and 11.

Conclusions

Selecting a worthy research question does not guarantee a worthwhile research project. The simplicity of the research circle presented in this chapter belies the complexity of the social research process. In the following chapters, I focus on particular aspects of the research process. Chapter 4 examines the interrelated processes of conceptualization and measurement, arguably the most important part of research. Measurement validity is the foundation for the other two aspects of validity. Chapter 5 reviews the meaning of generalizability and the sampling strategies that help us achieve this goal. Chapter 6 introduces causal validity and illustrates different methods for achieving it. Most of the remaining chapters then introduce different approaches to data collection—experiments, surveys, participant observation and intensive interviewing, evaluation research, comparative historical research, secondary data analysis, and content analysis—that help us, in different ways, achieve results that are valid.

Of course, our answers to research questions will never be complete or entirely certain. We always need to ground our research plans and results in the literature about related research. Our approach should be guided by explicit consideration of a larger theoretical framework. When we complete a research project, we should evaluate the confidence that can be placed in our conclusions, point out how the research could be extended, and consider the implications for social theory. Recall how the elaboration of knowledge about deterrence of domestic violence required sensitivity to research difficulties, careful weighing of the evidence, identification of unanswered questions, and consideration of alternative theories.

If you will conduct your own research, consider keeping a research journal in which you keep track of your decisions, question yourself at each step, and reflect on what you are reading and experiencing (Ravitch and Riggan 2016:216). It will help a lot when it comes time to write up the methods you have used.

Owning a large social science toolkit and even working with a good research team is no guarantee for making the right decisions about which tools to use and how to use them in the investigation of particular research problems, but you are now forewarned about, and thus I hope forearmed against, some of the problems that social scientists face in their work. I hope that you will return often to this chapter as you read the subsequent chapters, when you criticize the research literature, and when you design your own research projects. To be conscientious, thoughtful, and responsible—this is the mandate of every social scientist. If you formulate a feasible research problem, ask the right questions in advance, try to adhere to the research guidelines, and steer clear of the most common difficulties, you will be well along the road to fulfilling this mandate.

SAGE edge™

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Anomalous findings	53
Authenticity	58
Causal validity (internal validity)	58
Conflict theory	34
Cross-population generalizability (external validity)	59
Deductive research	50
Dependent variable	51
Direction of association	51
Empirical generalization	52
External validity (cross-population generalizability)	59
Functionalism	35
Generalizability	58
Hypothesis	51
Independent variable	51
Inductive research	53
Internal validity (causal validity)	58
Measurement validity	58
Normal science	35
Procedural justice theory	34
Rational choice theory	34
Replications	52
Research circle	50
Sample generalizability	59
Scientific paradigm	35
Serendipitous findings	53
Social research question	30
Symbolic interaction theory	34
Systematic review	44
Theory	33
Validity	57
Variable	51

Highlights

- Research questions should be feasible (within the time and resources available), socially important, and scientifically relevant.
- A theory is a logically interrelated set of propositions that helps us make sense of many interrelated phenomena and predict behavior or attitudes that are likely to occur when certain conditions are met.
- Building social theory is a major objective of social science research. Relevant theories should be

investigated before starting social research projects, and they should be used to focus attention on particular research questions and to draw out the implications of research findings.

- Rational choice theory focuses attention on the rational bases for social exchange and explains most social phenomena in terms of these motives.
- Symbolic interaction theory focuses attention on the meanings that people attach to and gain from social interaction and explains most social phenomena in terms of these meanings.
- Conflict theory focuses attention on the bases of conflict between social groups and uses these conflicts to explain most social phenomena.
- Functional theory explains social patterns in terms of their consequences for society as a whole and emphasizes the interdependence of social institutions and their common interest in maintaining the social order.
- A scientific paradigm is a set of beliefs that guide most scientific work in an area. Some researchers view positivism/postpositivism and interpretivism as alternative paradigms.
- Reviewing peer-reviewed journal articles that report prior research is an essential step in designing new research.
- The type of reasoning in most research can be described as primarily deductive or inductive. Research based on deductive reasoning proceeds from general ideas, deduces specific expectations from these ideas, and then tests the ideas with empirical data. Research based on inductive reasoning begins with specific data and then develops general ideas or theories to explain patterns in the data.
- It may be possible to explain unanticipated research findings after the fact, but such explanations have less credibility than those that have been tested with data collected for the purpose of the study.
- The scientific process can be represented as circular, with a path from theory to hypotheses, to data, and then to empirical generalizations. Research investigations may begin at different points along the research circle and traverse different portions of it. Deductive research begins at the point of theory, inductive research begins with data but ends with theory, and descriptive research begins with data and ends with empirical generalizations.
- Replications of a study are essential to establishing its generalizability in other situations. An ongoing line of research stemming from a particular research question should include a series of studies that, collectively, traverse the research circle multiple times.

Discussion Questions

1. Pick a social issue about which you think research is needed. Draft three research questions about this issue. Refine one of the questions and evaluate it in terms of the three criteria for good research questions.
2. Identify variables that are relevant to your three research questions. Now formulate three related hypotheses. Which are the independent and which are the dependent variables in these hypotheses?
3. If you were to design research about domestic violence, would you prefer an inductive approach or a deductive approach? Explain your preference. What would be the advantages and disadvantages of each approach? Consider in your answer the role of social theory, the value of searching the literature, and the goals of your research.
4. Sherman and Berk's (1984) study of the police response to domestic violence tested a prediction derived from rational choice theory. Propose hypotheses about the response to domestic violence that are consistent with conflict and symbolic interactionist theories. Which theory seems to you to provide the best framework for understanding domestic violence and how to respond to it?
5. Review my description of the research projects in the section "Types of Social Research" in Chapter 1. Can you identify the stages of each project corresponding to the points on the research circle? Did each project include each of the four stages? Which theory (or theories) seems applicable to each of these projects?
6. The research on battered women's help seeking used an exploratory research approach. Why do you think the researchers adopted this approach in these studies? Do you agree with their decision? Propose a research project that would address issues in one of these studies with a deductive approach.
7. Critique the Sechrist and Weil (2017) research on the police response to domestic violence from the standpoint of measurement validity, generalizability, and causal validity. What else would you like to know about this research so that you can strengthen your critique? What does consideration of the goal of authenticity add to your critique? How would you compare the strength of the research design and your confidence in the findings to the Sherman and Berk (1984) study?

Practice Exercises

1. Pair up with one other student and select one of the research articles available on the book's study site, at edge.sagepub.com/schutt9e. One of you should evaluate the research article in terms of its research strategy. Be generally negative but not unreasonable in your criticisms. The other student should critique the article in the same way but from a generally positive standpoint, defending its quality. Together, write a summary of the study's strong and weak points, or conduct a debate in class.
2. Research problems posed for explanatory studies must specify variables and hypotheses, which need to be stated properly and need to correctly imply any hypothesized causal relationship. The "Variables and Hypotheses" lessons, found in the interactive exercises on the study site, will help you learn how to do this.

   To use these lessons, choose one of the sets of "Variables and Hypotheses" exercises from the opening menu. About 10 hypotheses are presented in the lesson. After reading each hypothesis, name the dependent and independent variables and state the direction (positive or negative) of the relationship between them. In some of these interactive exercises, you must write in your own answer, so type carefully. The program will evaluate your answers. If an answer is correct, the program will present its version of the correct answer and go on to the next question. If you have made an error, the program will explain the error to you and give you another chance to respond. If your answer is unrecognizable, the program will instruct you to check your spelling and try again.
3. Now choose another article from the "Learning From Journal Articles" option on the study site. Read one article based on empirical research and diagram the process of research that it reports. Your diagram should have the structure of the research circle in Exhibit 2.11. How well does the process of research in this study seem to match the process symbolized in Exhibit 2.11? How much information is provided about each step in that process?
4. Review the section in this chapter on literature searching. Now choose a topic for investigation and search the social science literature for prior research on this topic. You will need to know how to use a database such as Sociological Abstracts at your own library as well as how to retrieve articles you locate (those that are available through your library). Try to narrow your search so that most of the articles you find are relevant to your topic (or broaden your search, if you don't find many relevant articles). Report your search terms and the results of your search with each term or combination of terms.

Ethics Questions

1. Sherman and Berk (1984) and those who replicated their research on the police response to domestic violence assigned persons accused of domestic violence by chance (randomly) to be arrested or not. The researchers' goal was to ensure that the people who were arrested were similar to those who were not arrested. Based on what you now know, do you feel that this random assignment procedure was ethical? Why or why not?
2. Concern with how research results are used is one of the hallmarks of ethical researchers, but deciding what form that concern should take is often difficult. You learned in this chapter about the controversy that occurred after Sherman and Berk (1984) encouraged police departments to adopt a pro-arrest policy in domestic abuse cases, based on findings from their Minneapolis study. Do you agree with the researchers' decision to suggest policy changes to police departments based on their study, in an effort to minimize domestic abuse? Several replication studies failed to confirm the Minneapolis findings. Does this influence your evaluation of what the researchers should have done after the Minneapolis study was completed?

Web Exercises

1. You can brush up on a range of social theorists at http://www.sociosite.net/topics/theory.php. Pick a theorist and read some of what you find. What social phenomena does this theorist focus on? What hypotheses seem consistent with his or her theorizing? Describe a hypothetical research project to test one of these hypotheses.
2. You've been assigned to write a paper on domestic violence and the law. To start, you can review relevant research on the American Bar Association's website at www.american bar.org/groups/domestic_violence/resources/statistics.html. What does the research summarized at this site suggest about the prevalence of domestic violence, its distribution among social groups, and its causes and effects? Write your answers in a one- to two-page report.

Video Interview Questions

Listen to the researcher interview for Chapter 2 at edge.sagepub.com/schutt9e.

1. What were the research questions that I focused on in the research project about homelessness and housing?
2. Why did we use a randomized experimental design?
3. I stated that the research design was consistent with reasonable ethical standards. Do you agree? Why or why not?
4. What were the answers to the two central research questions, as I described them?

To learn more, read Schutt (2011b), *Homelessness, Housing, and Mental Illness,* and pay particular attention to my appendix on research methods! http://www.hup.harvard.edu/catalog.php?isbn=9780674051010.

SPSS Exercises

1. Formulate four research questions about support for capital punishment—one question per research purpose: (1) exploratory, (2) descriptive, (3) explanatory, and (4) evaluative. Develop these questions so that you can answer at least two of them with variables in the GSS2016 data set you are using. Highlight these two.
2. Now, to develop some foundation from the literature, check the bibliography of this book for the following articles that drew on the GSS: Adalberto Aguirre Jr. and David Baker (1993); Steven Barkan and Steven Cohn (1994); Marian Borg (1997, 1998); Mark Warr (1995); and Robert Young (1992). How have social scientists used social theory to explain support for capital punishment? What potential influences on capital punishment have been tested? What influences could you test again with the 2016 GSS?
3. State four hypotheses in which support for capital punishment (CAPPUN) is the dependent variable and another variable in the GSS2016 data set is the independent variable. Justify each hypothesis in a sentence or two.
4. Test at least one hypothesis. Borg (1997) suggests that region might be expected to influence support for the death penalty. Test this as follows (after opening the GSS2016 file, as explained in Chapter 1, SPSS Exercise 3):
   a. Click on Analyze/Descriptive Statistics/Crosstabs.
   b. Highlight CAPPUN and click on the arrow so that it moves into the Rows box; highlight REGION and click on the arrow to move it into the Columns box.
   c. Click on Cells, click off Counts-Observed, and click on Percentages-Column.
   d. Click Continue and then OK. Inspect the table.
5. Does support for capital punishment vary by region? Scroll down to the percentage table (in which regions appear across the top) and compare the percentages in the Favor row for each region. Describe what you have found.
6. Now you can go on to test your other hypotheses in the same way, if you have the time. Because of space constraints, I can't give you more guidance, but I will warn you that there could be some problems at this point (e.g., if your independent variable has lots of values). Proceed with caution!

Developing a Research Proposal

Now you can prepare the foundation for your research proposal.

1. State a problem for research. If you have not already identified a problem for study, or if you need to evaluate whether your research problem is doable, a few suggestions should help get the ball rolling and keep it on course:
   a. Jot down questions that have puzzled you in some area having to do with people and social relations, perhaps questions that have come to mind while reading textbooks or research articles or even while hearing news stories. Don't hesitate to jot down many questions, and don't bore yourself—try to identify questions that really interest you.
   b. Now take stock of your interests, your opportunities, and the work of others. Which of your research questions no longer seem feasible or interesting? What additional research questions come to mind? Pick out a question that is of interest and seems feasible and that your other coursework suggests has been the focus of some prior research or theorizing.
   c. Write out your research question in one sentence, and elaborate on it in one paragraph. List at least three reasons why it is a good research question for you to investigate. Then present your proposal to your classmates and instructor for discussion and feedback.
2. Search the literature (and the web) on the research question you identified. Refer to the section on searching the literature for more guidance on conducting the search. Copy down at least 10

citations to articles (with abstracts from Sociological Abstracts or SocINDEX) and five websites reporting research that seems highly relevant to your research question; then look up at least five of these articles and three of the sites. Inspect the article bibliographies and the links on the website, and identify at least one more relevant article and website from each source.

3. Write a brief description of each article and website you consulted and evaluate its relevance to your research question. What additions or changes to your thoughts about the research question do the sources suggest?
4. Which general theoretical perspective do you believe is most appropriate to guide your proposed research? Write two paragraphs in which you (1) summarize the major tenets of the theoretical perspective you choose and (2) explain the relevance of this perspective to your research problem.
5. Propose at least two hypotheses that pertain to your research question. Justify these hypotheses in terms of the literature you have read.

Chapter 3 Research Ethics and Research Proposals

- Research That Matters, Questions That Count
- Historical Background
- Ethical Principles
    - Achievement of Valid Results
    - Honesty and Openness
    - Protection of Research Participants
        - *Avoid Harming Research Participants*
        - *Obtain Informed Consent*
        - *Avoid Deception in Research, Except in Limited Circumstances*
        - *Maintain Privacy and Confidentiality*
    - Consider Uses of Research So That Benefits Outweigh Risks
- The Institutional Review Board
- Research in the News: Some Social Scientists Are Tired of Asking for Permission
- Careers and Research
- Social Research Proposals
    - Case Study: Evaluating a Public Health Program
- Conclusions

Let's begin with a thought experiment (or a trip down memory lane, depending on your earlier exposure to this example). One spring morning as you are drinking coffee and reading the newspaper, you notice a small ad for a psychology experiment at the local university. "Earn money and learn about yourself," it says. Feeling a bit bored with your job as a high school teacher, you call and schedule an evening visit to the lab.

We Will Pay You $45 for One Hour of Your Time

Persons Needed for a Study of Memory

Research That Matters, Questions That Count

You are driving on the highway at about 3 p.m. on a Friday when you see a police officer standing by his squad car, lights flashing. The officer motions for you to pull off the road and stop in an area marked off with traffic cones. You are both relieved and surprised when someone in plain clothes working with the police officer then walks over to your car and asks if you would consent to be in a survey. You then notice two large signs that say NATIONAL ROADSIDE SURVEY and VOLUNTARY SURVEY. You are offered $10 to provide an oral fluid sample and answer a few additional questions on drug use.

This is what happened to 10,909 U.S. motorists between July 20 and December 1, 2007, at sites across the United States. Those who agreed to the oral fluid collection were also offered an additional $5 to complete a short alcohol and drug-use disorder questionnaire. Before they drove off, participants were also offered a $50 incentive for providing a blood sample. Drivers who were found to be too impaired to be able to drive safely (blood alcohol level above.05) were given a range of options, including switching with an unimpaired passenger, getting a free ride home, or spending a night in a local motel (at no expense to them). None were arrested or given citations and no crashes occurred in relation to the study. Those younger than 21 years and those who were pregnant were given informational brochures because of the special risk they face if they consume alcohol.

John H. Lacey and others from the Pacific Institute for Research and Evaluation, C. Debra Furr-Holden from Johns Hopkins University, and Amy Berning from the National Highway Traffic Safety Administration (NHTSA, which funded the study) reported the procedures for this survey in a 2011 article in the *Evaluation Review.* The article explained that all data collected were maintained as anonymous, so no research participants could be linked to their survey.

The 2007 National Roadside Survey identified 10.5% of the drivers as using illegal drugs and 3% as having taken medications.

What is your initial reaction to these research procedures, involving collaboration with the police, diversion of drivers, and measurement of substance abuse?

1. The institute's institutional review board (IRB) reviewed all staff training and operational procedures and a human subjects protection training module was used to prepare interviewers for the roadside encounters. Do you think human subjects were protected? How about the procedures with impaired drivers?
2. Do you think that the potential benefits of this study for improving policies about impaired driving would outweigh concerns about interference with individuals' activities?

In this chapter, you will learn about standards and procedures for the protection of human subjects in research. By the end of the chapter, you will have a much firmer basis for answering the questions I have posed. After you finish the chapter, test yourself by reading the 2011 *Evaluation Review* article at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 3 at edge.sagepub.com/schutt9e.

Lacey, John H., Tara Kelley-Baker, Robert B. Voas, Eduardo Romano, C. Debra Furr-Holden, Pedro Torres, and Amy Berning. 2011. "Alcohol- and Drug-Involved Driving in the United States: Methodology for the 2007 National Roadside Survey." *Evaluation Review* 35:319–353.

You arrive at the assigned room at the university, ready for an interesting hour or so, and are impressed immediately by the elegance of the building and the professional appearance of the personnel. In the waiting room, a man dressed in a lab coat turns and introduces himself and explains that as a psychologist, he is interested in the question of whether people learn things better when they are punished for making a mistake. He then explains that his experiment on punishment and learning will help answer this question.

"The experimenter" [as we'll refer to him from now on] says he will write either *teacher* or *learner* on small identical slips of paper and then asks you and another person in the room to draw out one. Yours says teacher.

The experimenter now says, in a matter-of-fact way, "All right. Now the first thing we'll have to do is to set the learner up so that he can get some type of punishment."

He leads you both behind a curtain, sits the learner down, attaches a wire to his left wrist, and straps both his arms to the chair so that he cannot remove the wire (see Exhibit 3.1). The wire is connected to a console with 30 switches and a large dial on the other side of the room. The experimenter asks you to hold the end of the wire, walks back to the control console, flips several switches, and you hear a clicking noise. The dial moves and you feel an electric shock. The experimenter explains that the machine is calibrated so that it will not cause permanent injury, but acknowledges that when it is turned up all the way it is very, very painful and can result in severe, although momentary, discomfort.

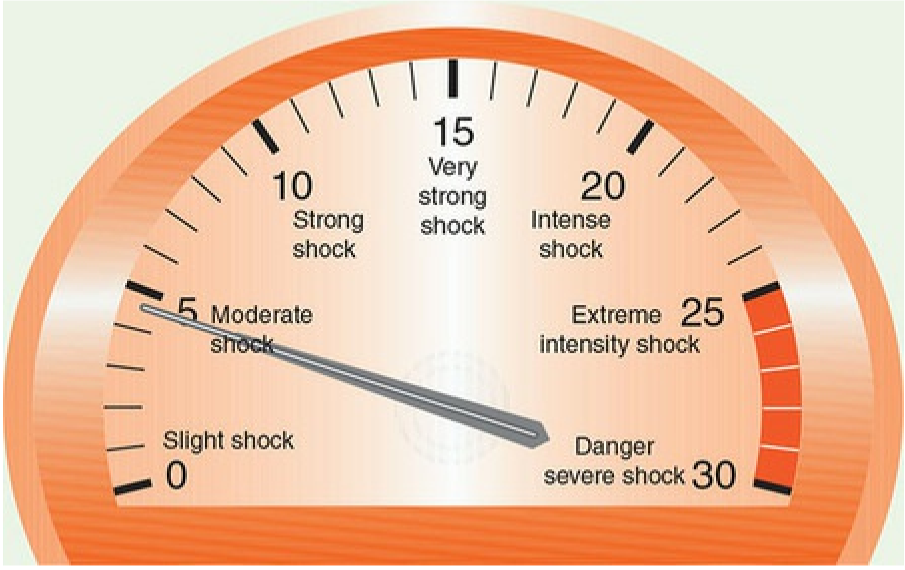
Now you walk back to the other side of the room (so that the learner is behind the curtain) and sit before the console. The experimental procedure is simple: (1) You read aloud a series of word pairs, such as *blue box, nice day, wild duck,* and so on. (2) You then read one of the first words from those pairs and a set of four words, one of which contains the original paired word. (3) If the learner states the word that he thinks was paired with the first word you read, you are to compliment him and move on to the next word. If he makes a mistake, you flip a switch on the console and the learner feels a shock. (4) After each mistake, you are to flip the next switch on the console, progressing from a mark labeled *slight shock,* on up to marks labeled *intense shock, extreme intensity shock,* and *danger: severe shock.*

You begin. As you turn the dial, the learner's responses increase in intensity from a grunt at the tenth mark (strong shock) to painful groans at higher levels, anguished cries to "get me out of here" at the extreme intensity shock levels, to a deathly silence at the highest level. If you indicate discomfort at administering the stronger shocks, the experimenter tells you, "The experiment requires that you continue." Now, please note on the meter in Exhibit 3.2 the most severe shock that you would agree to give to the learner.

Exhibit 3.1 Learner Strapped in Chair With Electrodes

Source: From the film OBEDIENCE. Copyright © 1968 by Stanley Milgram, copyright renewed 1993 by Alexandra Milgram and distributed by Alexander Street Press.

Exhibit 3.2 Shock Meter


0
Slight shock
5 Moderate shock
10 Strong shock
15 Very strong shock
20 Intense shock
25 Extreme intensity shock
30 Danger severe shock

You may very well recognize that this thought experiment is a slightly simplified version of Milgram's obedience experiments, begun at Yale University in 1960. The 40 New Haven adults who volunteered for the experiment administered an average level of shock of 24.53, or a level of about extreme intensity shock and close to danger: severe shock. Almost two-thirds complied with the experimenter's demands all the way to the top of the scale (originally labeled simply as *XXX*).

There is abundant evidence from the subjects' own observed high stress and their subsequent reports that many subjects really believed that the learner was receiving actual, hurtful shocks. Are you surprised by the subjects' responses? Do you think the results of this experiment tell us about how people behave in the real world? Can you imagine why Milgram's (1963) research on obedience ultimately had as profound an influence on the way social scientists think about research ethics as it had on the way they understand obedience to authority?

Historical Background

Concern with ethical practice in relation to people who are in some respect dependent, whether as patients or research subjects, is not a new idea. Ethical guidelines for medicine trace back to Hippocrates in 5 BC Greece (Hippocratic Oath, n.d.), and the American Medical Association (AMA) adopted the world's first formal professional ethics code in medicine in 1847 (AMA 2011). Current AMA ethical principles include respecting patient rights, maintaining confidentiality, and regarding "responsibility to the patient as paramount" (AMA 2011). Yet the history of medical practice makes it clear that having an ethics code is not sufficient to ensure ethical practice, at least when there are clear incentives to do otherwise.

A defining event occurred in 1946, when the Nuremberg War Crime Trials exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science." In 1961, Stanley's Milgram's research on obedience also generated controversy about participant protections (Perry 2013:37). In 1966, Harvard medical researcher Henry K. Beecher published an article in the prestigious *New England Journal of Medicine* that described 22 unethical experimental studies of human subjects, usually without their knowledge (Israel 2014:32–33). In 1970, sociologist Laud Humphreys's book, *Tearoom Trade,* revealed his secretive observations of men engaged in sex with other men in public bathrooms, followed by his interviewing them in their homes under false pretenses (described later in this chapter); and in 1971, sociologist Philip Zimbardo conducted his famous prison simulation experiment at Stanford University to learn how ordinary people would treat each other if they took roles of being prisoners and guards (the answer: not well) (Israel 2014:24–28).

Milgram's obedience experiments: Experiments begun in 1960 at Yale University by psychologist Stanley Milgram to determine the likelihood of people following orders from an authority despite their own sentiments; widely cited as helping to understand the emergence of phenomena such as Nazism and mass cults.

Nuremberg War Crime Trials: The International Military Tribunal held by the victorious Allies after World War II in Nuremberg, Germany, that exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science."

In 1972, Americans learned from news reports that researchers funded by the U.S. Public Health Service had followed 600 low-income African American men in the Tuskegee Study of Untreated Syphilis in the Negro Male since the 1930s, collecting data to study the "natural" course of the illness (see Exhibit 3.3) (National Center for Bioethics in Research and Health Care n.d.). There was no effective treatment for the disease when the study began and participants received free medical exams, meals, and burial insurance.

However, the study was deceptive from its start: The men were told they were being treated for "bad blood," whether they had syphilis or not (399 did), and were not asked for their consent to be studied. What made the Tuskegee Study so shocking was that many participants were not informed of their illness and were not offered treatment with penicillin after the drug was recognized as effective in 1945 and came into large-scale use by 1947. The research was ended only after the study was exposed. Congressional hearings began in 1973, and an out-of-court settlement of almost $10 million was reached in 1974. It was not until 1997 that President Bill Clinton issued an official public apology on behalf of the U.S. government (CDC 2009).

Exhibit 3.3 Tuskegee Syphilis Experiment

Source: Tuskegee Syphilis Study Administrative Records. Records of the Centers for Disease Control and Prevention. National Archives—Southeast Region (Atlanta).

These and other widely publicized cases convinced many observers that formal review procedures were needed to protect research participants. As concerns about research ethics increased, the U.S. government created a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research and charged it with developing guidelines (Kitchener and Kitchener 2009:7). The commission's 1979 Belmont Report (Department of Health, Education, and Welfare 1979) established three basic ethical principles for the protection of human subjects:

- Respect for persons: treating persons as autonomous agents and protecting those

with diminished autonomy
- Beneficence: minimizing possible harms and maximizing benefits
- Justice: distributing benefits and risks of research fairly

Tuskegee Study of Untreated Syphilis in the Negro Male: The U.S. Public Health Service study of the "natural" course of syphilis that followed 399 low-income African American men from the 1930s to 1972, without providing them with penicillin after the drug was discovered to treat the illness. The study was stopped after it was exposed in 1972, resulting in an out-of-court settlement and then, in 1997, an official public apology by President Bill Clinton.

Belmont Report: Guidelines developed by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979 for the protection of human subjects.

Respect for persons: The ethical principle of treating persons as autonomous agents and protecting those with diminished autonomy in research involving human subjects that was included in the *Belmont Report.*

Beneficence: The ethical requirement of minimizing possible harms and maximizing benefits in research involving human subjects that was included in the *Belmont Report.*

Justice: The ethical principle of distributing benefits and risks of research fairly in research involving human subjects that was included in the *Belmont Report.*

The U.S. Department of Health and Human Services (DHHS) and the U.S. Food and Drug Administration then translated these principles into specific regulations that were adopted in 1991 as the Federal Policy for the Protection of Human Subjects, also known as the Common Rule (Title 45 of Code of Federal Regulations [CFR], Part 46). This policy has shaped the course of social science research ever since, by requiring organizations that sponsor federally funded research—including universities—to establish committees that review all research proposed at the institution and ensure compliance with the federal human subjects requirements when the research is conducted. Professional associations such as the American Sociological Association (ASA), university review boards, and ethics committees in other organizations also set standards for the treatment of human subjects by their members, employees, and students, although these standards are all designed to comply with the federal policy.

After 25 years, the Common Rule was revised, with modifications announced in January 2017 with an implementation date of January 2018 (Menikoff et al. 2017). The revisions relaxed some requirements for social science research and made several other important changes that are consequential for both medical and social science researchers. Although their implementation may be delayed until 2019, these regulations, as revised, inform the discussion that follows (Baumann 2017; Chadwick 2017).

Ethical Principles

The ASA, like other professional social science organizations, has adopted, for practicing sociologists, ethical guidelines that are more specific than the federal regulations. Professional organizations may also review complaints of unethical practices when asked.

The *Code of Ethics* of the ASA (1999) is summarized on the ASA website (http://www.asanet.org/membership/code-ethics); the complete text of the code is also available at this site. The general principles articulated in the code are intended to guide professional practice in diverse settings, while some of the specific standards focus specifically on the protection of human subjects in research.

According to the general principles, sociologists should be committed in their work to high levels of competence, to practicing with integrity, and to maintaining responsibility for their actions. They must also respect the rights, dignity, and diversity of others, including research participants, as well as be socially responsible to their communities and use research to contribute to the public good. The following sections discuss the most important implications of these principles for the conduct of research, including the protection of human subjects.

Achievement of Valid Results

Commitment to achieving valid results is the necessary starting point for ethical research practice. Simply put, we have no business asking people to answer questions, submit to observations, or participate in experimental procedures if we are simply seeking to verify our preexisting prejudices or convince others to take action on behalf of our personal interests. The pursuit of objective knowledge about human behavior—the goal of validity—motivates and justifies our investigations and gives us some claim to the right to influence others to participate in our research. Knowledge is the foundation of human progress as well as the basis for our expectation that we, as social scientists, can help people achieve a brighter future. If we approach our research projects objectively, setting aside our personal predilections in the service of learning a bit more about human behavior, we can honestly represent our actions as potentially contributing to the advancement of knowledge.

Federal Policy for the Protection of Human Subjects (also known as the Common Rule): Specific regulations adopted in 1991 by the U.S. Department of Health and Human Services and the U.S. Food and Drug Administration that were based on the principles of the *Belmont Report.* Revised in 2017, with new requirements in effect, January 2018.

Code of Ethics: The professional code of the American Sociological Association for the treatment of human subjects by members, employees, and students; designed to comply with federal policy and revised in 1997.

Milgram made a strong case in his 1963 article and 1974 book on the obedience experiments that he was committed to achieving valid results—to learning how and why obedience influences behavior. He tied his motivations directly to the horror of the Holocaust, to which the world's attention had been drawn once again by the capture and trial of Adolf Hitler's mastermind of that genocide, Adolf Eichmann (Perry 2013:210). In Milgram's (1963) own words,

> It has been reliably established that from 1933–45 millions of innocent persons were systematically slaughtered on command. . . . Obedience is the psychological mechanism that links individual action to political purpose. It is the dispositional cement that binds men to systems of authority . . . for many persons obedience may be a deeply ingrained behavior tendency. . . . Obedience may [also] be ennobling and educative and refer to acts of charity and kindness, as well as to destruction. (p. 371)

Milgram (1963) then explains how he devised experiments to study the process of obedience in a way that would seem realistic to the subjects and still allow "important variables to be manipulated at several points in the experiment" (p. 372). According to Milgram, every step in the experiment was carefully designed to ensure that subjects received identical stimuli and that their responses were measured carefully. The experiment's design also reflected what had become in the preceding 30 years a tradition in social psychology of laboratory experiments that used deception to create different believable conditions for participants (Perry 2013:31–35).

Milgram (1963:377) made every effort to convince readers that "the particular conditions" of his experiment created the conditions for achieving valid results. These particular conditions included the setting for the experiment at Yale University, its purported "worthy purpose" to advance knowledge about learning and memory, and the voluntary participation of the subject as well as of the learner—as far as the subject knew. Milgram then tested the importance of some of these "particular conditions" (e.g., the location at Yale) in replications of the basic experiment (Milgram 1965).

However, not all social scientists agreed that Milgram's approach could achieve valid results. Milgram's first article on the research, "Behavioral Study of Obedience," was published in 1963 in the *Journal of Abnormal and Social Psychology.* In the next year, the *American Psychologist* published a critique of the experiment's methods and ethics by the psychologist Diana Baumrind (1964). Her critique begins with a rejection of the external validity—the generalizability—of the experiment, because

> The laboratory is unfamiliar as a setting and the rules of behavior ambiguous. . . . Therefore, the laboratory is not the place to study degree of obedience or suggestibility, as a function of a particular experimental condition. [And so,] the parallel between authority-subordinate relationships in Hitler's Germany and in Milgram's laboratory is unclear. (pp. 421–423)

Milgram (1964) quickly published a rejoinder in which he disagreed with (among other things) the notion that it is inappropriate to study obedience in a laboratory setting: "A subject's obedience is no less problematical because it occurs within a social institution called the psychological experiment" (p. 850).

Milgram (1974:169–178) also argued in his later book that his experiment had been replicated in other places and settings with the same results, that there was considerable evidence that the subjects had believed that they actually were administering shocks, and that the "essence" of his experimental manipulation—the request that subjects comply with a legitimate authority—was also found in the dilemma faced by people in Nazi Germany, soldiers at the My Lai massacre in Vietnam, and even the cultists who drank poison in

Jonestown, Guyana, at the command of their leader, Jim Jones (Miller 1986:182–183).

Baumrind (1985) was still not convinced. In a follow-up article in *American Psychologist,* she argued that "far from illuminating real life, as he claimed, Milgram in fact appeared to have constructed a set of conditions so internally inconsistent that they could not occur in real life" (p. 171). Although Milgram died in 1984, the controversy did not. A recent review of the transcripts and interviews with many participants raises additional concerns about the experiment's validity (Perry 2013). Milgram understated the "experimenter's" efforts to get the subjects to comply, he overstated the subjects' level of obedience, he never publicized one condition in which most subjects refused to give strong shocks when the "learner" was a friend, and he didn't acknowledge that even those classified as "obedient" were looking for a way to get out of the experiment. His claim that the results were replicated in similar experiments around the world was only partially true, and it seems clear from the transcripts and interviews that the aura created by the location at Yale University and the emphasis on a contribution to "science" influenced many participants.

Do you agree with Milgram's assumption that obedience could fruitfully be studied in the laboratory? Do you find merit in Baumrind's criticism? Are you troubled by the new evidence that Milgram may have presented his evidence selectively, to make his conclusions as convincing as possible? Will your evaluation of the ethics of Milgram's experiments be influenced by your answers to these questions? Should our ethical judgments differ depending on whether we decide that a study provides valid information about important social psychological processes? Should it matter that a 2005 replication of Milgram's experiment (with less severe "shocks") for ABC TV supported Milgram's conclusions (Perry 2013:275–279)?

I can't answer these questions for you, but before you dismiss them as inappropriate when we are dealing with ethical standards for the treatment of human subjects, bear in mind that both Milgram and his strongest critic at the time, Baumrind, buttressed their ethical arguments with assertions about the validity (or invalidity) of the experimental results. It is hard to justify *any* risk for human subjects, or even *any* expenditure of time and resources, if our findings tell us nothing about human behavior.

Honesty and Openness

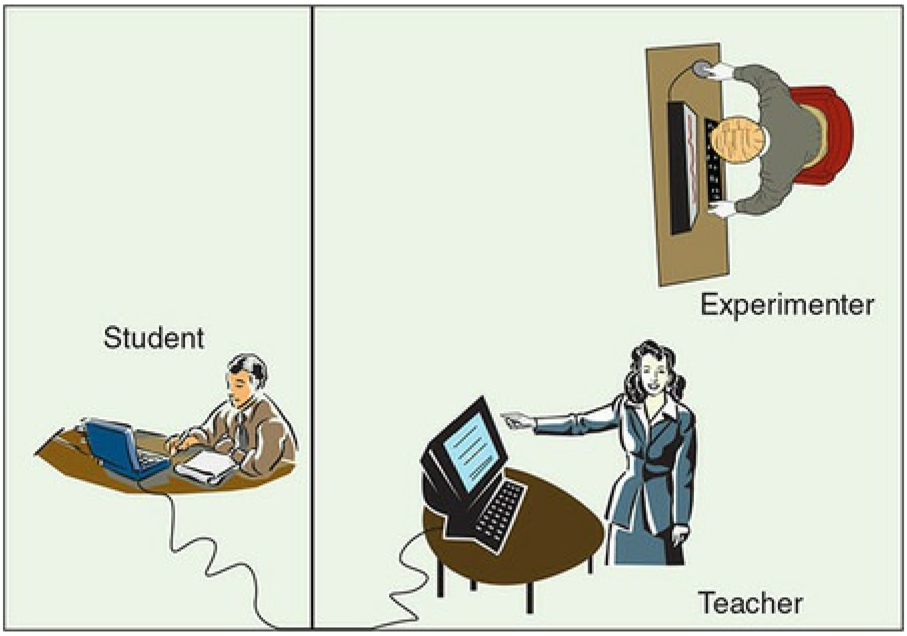
The scientific concern with validity requires, in turn, that scientists be open in disclosing their methods and honest in presenting their findings. In contrast, research distorted by political or personal pressures to find particular outcomes or to achieve the most marketable results is unlikely to be carried out in an honest and open fashion. To assess the validity of a researcher's conclusions and the ethics of their procedures, you need to know exactly how the research was conducted. This means that articles or other reports must include a detailed methodology section, perhaps supplemented by appendixes containing the research instruments, or websites or an address where more information can be obtained.

Milgram presented his research in a way that would signal his adherence to the goal of honesty and openness. His initial 1963 article included a detailed description of study procedures, including the text of the general introduction to participants, the procedures involved in the learning task—the "shock generator," the administration of the "sample shock," the shock instructions and the preliminary practice run, the standardized feedback from the "victim" and from the experimenter—and the measures used. Many more details, including pictures, were provided in Milgram's (1974) subsequent book (see Exhibit 3.4).

The act of publication itself is a vital element in maintaining openness and honesty. Others can review and question study procedures and so generate an open dialogue with the researcher. Although Milgram disagreed sharply with Baumrind's criticisms of his experiments, their mutual commitment to public discourse in journals widely available to social scientists resulted in a more comprehensive presentation of study procedures and a more thoughtful discourse about research ethics. Almost 50 years later, this commentary continues to inform debates about research ethics (Cave and Holm 2003).

The latest significant publication in this open dialogue about Milgram's work actually challenges his own commitment to the standard of openness and honesty. Gina Perry's (2013) *Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments* reveals many misleading statements by Milgram about participants' postexperiment debriefing, about adherence to the treatment protocol, about the extent of participants' apparent distress, and about the extent of support for his favored outcome.

Exhibit 3.4 Diagram of Milgram Experiment



Source: Northern Illinois University Department of Psychology.

Openness about research procedures and results thus goes hand in hand with honesty in research design and in research reporting. Despite this need for openness, some researchers may hesitate to disclose their procedures or results to prevent others from building on their ideas and taking some of the credit or, as may have occurred with Milgram, to make their procedures seem more acceptable or their findings more impressive. We just can't assume that everyone conducting research will avoid cutting corners in their work or failing to explain their procedures fully, or even fabricating evidence if it helps to achieve their career goals. Just in 2011, Harvard psychology professor Marc Hauser resigned after an investigation concluded that data he had reported in scientific publications was fabricated (Israel 2014:155).

Conflicts of interest may occur when a researcher has a significant financial stake in the design or outcome of the research (American Sociological Association n.d.). Receiving speaking fees, consulting fees, patents or royalties, and other financial benefits from organizations that could be affected by research conclusions may lead researchers to distort their decisions and findings so as to protect these financial benefits. Many medical schools have regulations that prohibit their staff (including researchers) from accepting any tangible benefits from pharmaceutical companies for this reason. Both federal research funding agencies and journal editors require disclosure of possible conflicts of interest so that others can scrutinize the extent to which these conflicts may have lessened researchers' honesty and openness (Fisher and Anushko 2008:96–97). Unfortunately, experimental research suggests that disclosure does not reduce trust in advice from people who have disclosed a conflict of interest (Humphries 2011:K3).

Conflict of interest: When a researcher has a significant financial stake in the design or outcome of his or her own research.

Protection of Research Participants

Several standards concerning the treatment of human subjects are emphasized in federal regulations and the ethical guidelines adopted by many professional social science organizations:

- Research should cause no harm to subjects.
- Participation in research should be voluntary; therefore, subjects must give their informed consent to participate in the research and researchers must disclose their identity.
- Researchers should avoid deception, except in limited circumstances.
- Anonymity or confidentiality must be maintained for individual research participants unless it is waived voluntarily and explicitly.
- Scientists must consider the uses of a research project so that its benefits outweigh any foreseeable risks.

Each of these standards became a focus of debate about Milgram's experiments, so we will return frequently to that debate to keep our discussion realistic. We will also refer frequently to the ASA code to keep our treatment current. You will soon realize that there is no simple answer to the question: What is (or isn't) ethical research practice? The issues are just too complicated and the relevant principles too subject to different interpretations. But, I do promise that by the time you finish this chapter, you will be aware of the major issues in research ethics and be able to make informed, defensible decisions about the ethical conduct of social science research.

Avoid Harming Research Participants

Although this standard may seem straightforward, it can be difficult to interpret in specific cases and harder yet to define in a way agreeable to all social scientists. Does it mean that subjects should not be harmed psychologically as well as physically at all? That they should feel no anxiety or distress whatsoever during the study or only after their involvement ends? Should the possibility of any harm, no matter how remote, deter research?

Before we address these questions with respect to Milgram's experiments, a verbatim transcript of one session will give you an idea of what participants experienced (Milgram 1965:67):

150 volts delivered. You want me to keep going?

165 volts delivered. That guy is hollering in there. There's a lot of them here. He's liable to have a heart condition. You want me to go on?

180 volts delivered. He can't stand it! I'm not going to kill that man in there! You hear him hollering? He's hollering. He can't stand it. . . . I mean who is going to take responsibility if anything happens to that gentleman? *[The experimenter accepts responsibility.]* All right.

195 volts delivered. You see he's hollering. Hear that. Gee, I don't know. *[The experimenter says: "The experiment requires that you go on."]* I know it does, sir, but I mean—hugh—he don't know what he's in for. He's up to 195 volts.

210 volts delivered.

225 volts delivered.

240 volts delivered.

This experimental manipulation generated "extraordinary tension" (Milgram 1963:377):

> Subjects were observed to sweat, tremble, stutter, bite their lips, groan and dig their fingernails into their flesh. . . . Full-blown, uncontrollable seizures were observed for 3 subjects. One . . . seizure so violently convulsive that it was necessary to call a halt to the experiment [for that individual]. (p. 375)

An observer (behind a one-way mirror) reported, "I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse" (Milgram 1963:377).

From critic Baumrind's (1964:422) perspective, this emotional disturbance in subjects was "potentially harmful because it could easily effect an alteration in the subject's self-image or ability to trust adult authorities in the future." Milgram (1964) quickly countered,

> Momentary excitement is not the same as harm. As the experiment progressed there was no indication of injurious effects in the subjects; and as the subjects themselves strongly endorsed the experiment, the judgment I made was to continue the experiment. (p. 849)

When Milgram (1964:849) surveyed participants in a follow-up, 83.7% endorsed the statement that they were "very glad" or "glad" "to have been in the experiment," 15.1% were "neither sorry nor glad," and just 1.3% were "sorry" or "very sorry" to have participated (p. 849). Interviews by a psychiatrist a year later found no evidence "of any traumatic reactions" (p. 849)—although he did not disclose that of 780 initial participants, only 140 were invited for an interview and only 32 of those accepted the invitation (Perry 2013:217). After these later revelations, Milgram's (1977:21) subsequent argument that "the central moral justification for allowing my experiment is that it was judged acceptable by those who took part in it" rings hollow.

Milgram (1963) also reported that he attempted to minimize harm to subjects with postexperimental procedures "to assure that the subject would leave the laboratory in a state of well being" (p. 374). He said that a friendly reconciliation was arranged between the subject and the victim, and an effort was made to reduce any tensions that arose as a result of the experiment, but it turns out that his "dehoaxing" was typically very brief and did not disclose the deception to most participants. Most participants did not receive a letter informing them of the nature of the experiment until almost a year had passed (Milgram 1964:849; Perry 2013:72, 84).

Baumrind (1964:422) was unconvinced even without knowing of these later revelations: "It would be interesting to know what sort of procedures could dissipate the type of emotional disturbance just described [citing Milgram 1964]."

In a later article, Baumrind (1985:168) dismissed the value of the self-reported "lack of harm" to subjects who had been willing to participate in the experiment—although noting that still 16% did *not* endorse the statement that they were "glad" they had participated in the experiment. Baumrind (1985:169) also argued that research indicates most introductory psychology students (and some students in other social sciences) who have participated in a deception experiment report a decreased trust in authorities as a result—a tangible harm in itself.

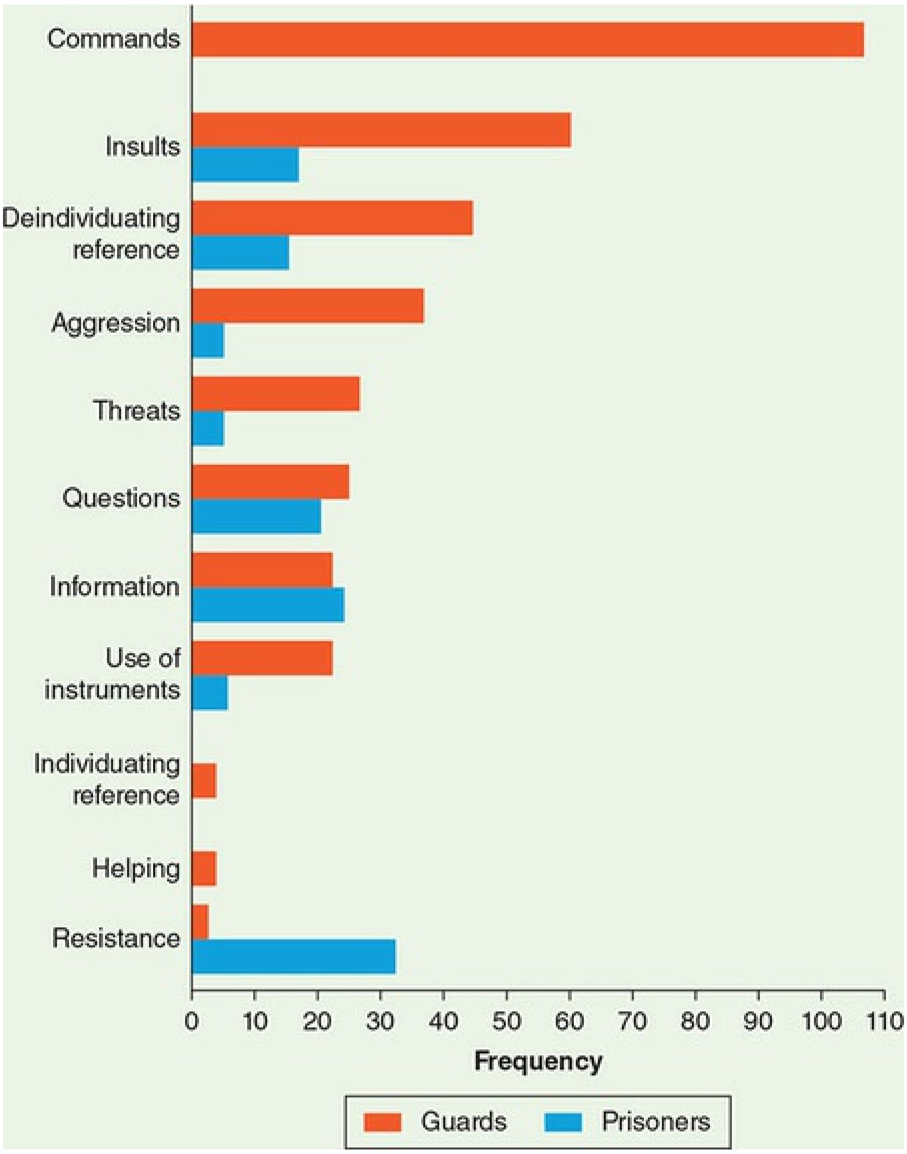
Many social scientists, ethicists, and others concluded that Milgram's procedures had not harmed the subjects and so were justified for the knowledge they produced, but others sided with Baumrind's criticisms (Miller 1986:88–138; Perry 2013:269). Perry's (2013:77–78) recent investigation found even more evidence of psychological harm, including feelings of shame that had persisted since the experiment. The experimental records also reveal that debriefing never occurred for some participants and was very limited for almost all (Perry 2013:76–84). Most were not told after the experiment that the shocks were fake; the usual "dehoaxing" consisted of the "learner" reassuring the "teacher" that the shocks he had received were not harmful.

What is your opinion of the possibility for harm at this point? Does Milgram's debriefing process relieve your concerns? Are you as persuaded by the subjects' own endorsement of the experiment as was Milgram?

What about possible harm to the subjects of the famous Zimbardo prison simulation study at Stanford University (Haney, Banks, and Zimbardo 1973)? The study was designed to investigate the impact of social position on behavior—specifically, the impact of being either a guard or a prisoner in a prison, a “total institution.” The researchers selected apparently stable and mature young male volunteers and asked them to sign a contract to work for 2 weeks as a guard or a prisoner in a simulated prison. Within the first 2 days after the prisoners were incarcerated by the “guards” in a makeshift basement prison, the prisoners began to be passive and disorganized, while the guards became “sadistic”—verbally and physically aggressive (see Exhibit 3.5). Five “prisoners” were soon released for depression, uncontrollable crying, fits of rage, and, in one case, a psychosomatic rash. Instead of letting things continue for 2 weeks as planned, Philip Zimbardo and his colleagues terminated the experiment after 6 days to avoid harming the subjects. Feelings of stress among the participants who played the role of prisoner seemed to be relieved by discussions in special postexperiment encounter sessions, while follow-up during the next year indicated no lasting negative effects on the participants and some gains in self-insight.

Zimbardo prison simulation study: The famous prison simulation study at Stanford University by psychologist Philip Zimbardo, designed to investigate the impact of social position on behavior—specifically, the impact of being either a guard or a prisoner in a “total institution”; widely cited as demonstrating the likelihood of emergence of sadistic behavior in guards.

Exhibit 3.5 Chart of Guard and Prisoner Behavior



Source: Adapted from The Lucifer Effect: Understanding How Good People Turn Evil by Philip G. Zimbardo. Copyright 2007 by Philip G. Zimbardo, Inc. Used by permission of Random House, Inc., an imprint and division of Penguin Random House LLC; and Random House Group Ltd. All rights reserved.

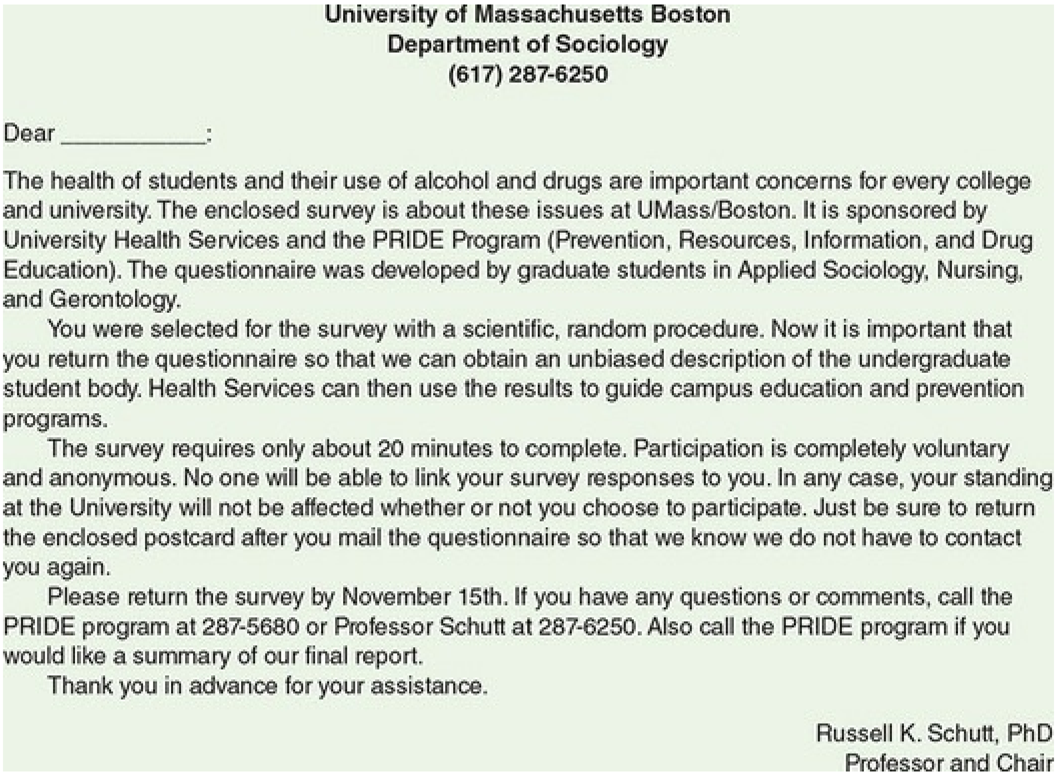
Would you ban such experiments because of the potential for harm to subjects? Does it make any difference to you that Zimbardo's and Milgram's experiments seemed to yield significant insights into the effect of a social situation on human behavior—insights that could be used to improve prisons or perhaps lessen the likelihood of another holocaust (Reynolds 1979:133–139)? Do you believe that this benefit outweighs the foreseeable risks? Arthur Miller (1986) argued that real harm "could result from *not doing* research on destructive obedience" (p. 138) and other troubling human behaviors. What if the researchers themselves—as was true of both Milgram (1974:27–31) and Zimbardo (2007)—did not foresee the potential harm to their participants?

Obtain Informed Consent

The requirement of informed consent is also more difficult to define than it first appears. To be informed, consent must be given by the persons who are competent to consent, have consented voluntarily, are fully informed about the research and know who is conducting the research, and have comprehended what they have been told (Reynolds 1979). Yet you probably realize, as Baumrind (1985) did, that because of the inability to communicate perfectly, "full disclosure of everything that could possibly affect a given subject's decision to participate is not possible, and therefore cannot be ethically required" (p. 165).

Obtaining informed consent creates additional challenges for researchers. The researcher's actions and body language should help convey his or her verbal assurance that consent is voluntary. The language of the consent form must be clear and understandable to the research participants and yet sufficiently long and detailed to explain what will actually happen in the research. Consent Forms A (Exhibit 3.6) and B (Exhibit 3.7) illustrate two different approaches to these trade-offs.

Exhibit 3.6 Consent Form A



University of Massachusetts Boston
Department of Sociology
(617) 287-6250

Dear ___________:

The health of students and their use of alcohol and drugs are important concerns for every college and university. The enclosed survey is about these issues at UMass/Boston. It is sponsored by University Health Services and the PRIDE Program (Prevention, Resources, Information, and Drug Education). The questionnaire was developed by graduate students in Applied Sociology, Nursing, and Gerontology.

You were selected for the survey with a scientific, random procedure. Now it is important that you return the questionnaire so that we can obtain an unbiased description of the undergraduate student body. Health Services can then use the results to guide campus education and prevention programs.

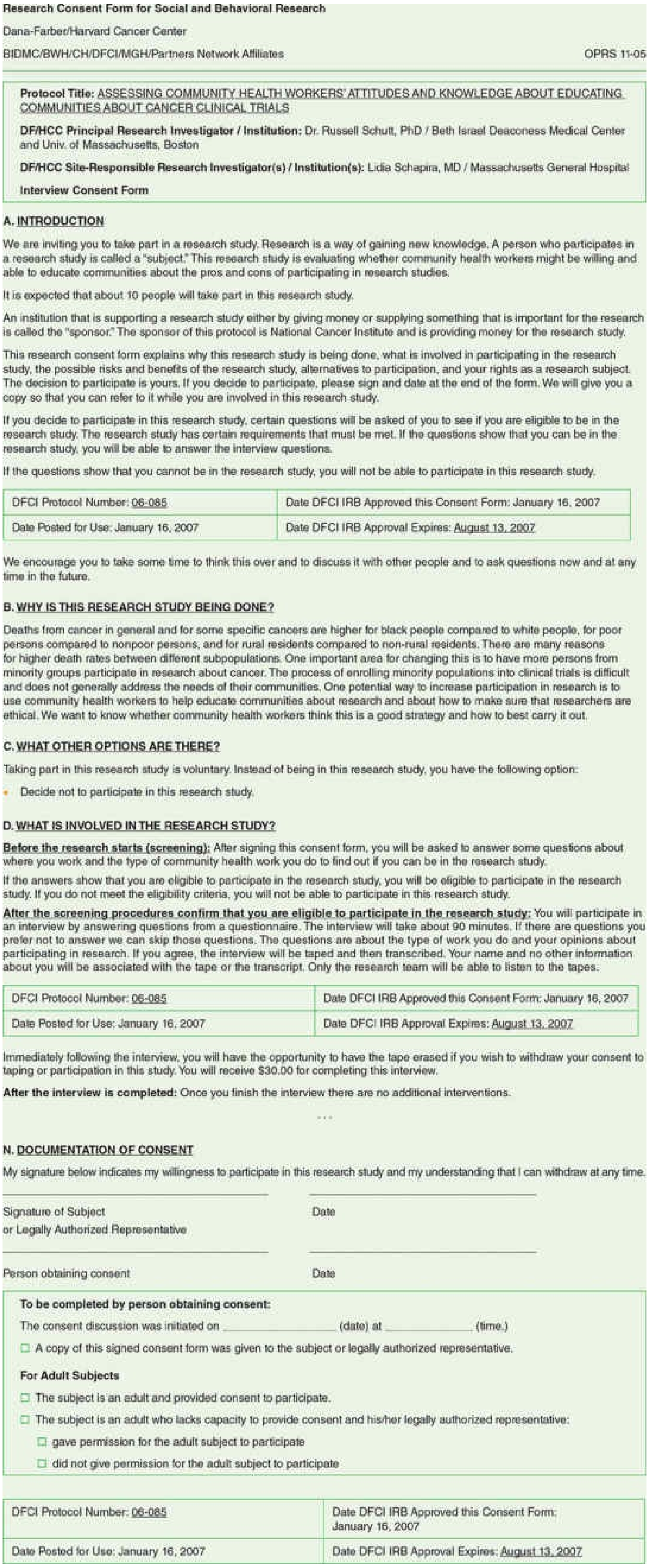
The survey requires only about 20 minutes to complete. Participation is completely voluntary and anonymous. No one will be able to link your survey responses to you. In any case, your standing at the University will not be affected whether or not you choose to participate. Just be sure to return the enclosed postcard after you mail the questionnaire so that we know we do not have to contact you again.

Please return the survey by November 15th. If you have any questions or comments, call the PRIDE program at 287-5680 or Professor Schutt at 287-6250. Also call the PRIDE program if you would like a summary of our final report.

Thank you in advance for your assistance.

Russell K. Schutt, PhD
Professor and Chair

Exhibit 3.7 Consent Form B



Research Consent Form for Social and Behavioral Research

Dana-Farber/Harvard Cancer Center

BIDMC/BWH/CH/DFCI/MGH/Partners Network Affiliates OPRS 11-05

Protocol Title: ASSESSING COMMUNITY HEALTH WORKERS' ATTITUDES AND KNOWLEDGE ABOUT EDUCATING COMMUNITIES ABOUT CANCER CLINICAL TRIALS

DF/HCC Principal Research Investigator / Institution: Dr. Russell Schutt, PhD / Beth Israel Deaconess Medical Center and Univ. of Massachusetts, Boston

DF/HCC Site-Responsible Research Investigator(s) / Institution(s): Lidia Schapira, MD / Massachusetts General Hospital

Interview Consent Form

A. INTRODUCTION

We are inviting you to take part in a research study. Research is a way of gaining new knowledge. A person who participates in a research study is called a "subject." This research study is evaluating whether community health workers might be willing and able to educate communities about the pros and cons of participating in research studies.

It is expected that about 10 people will take part in this research study.

An institution that is supporting a research study either by giving money or supplying something that is important for the research is called the "sponsor." The sponsor of this protocol is National Cancer Institute and is providing money for the research study.

This research consent form explains why this research study is being done, what is involved in participating in the research study, the possible risks and benefits of the research study, alternatives to participation, and your rights as a research subject. The decision to participate is yours. If you decide to participate, please sign and date at the end of the form. We will give you a copy so that you can refer to it while you are involved in this research study.

If you decide to participate in this research study, certain questions will be asked of you to see if you are eligible to be in the research study. The research study has certain requirements that must be met. If the questions show that you can be in the research study, you will be able to answer the interview questions.

If the questions show that you cannot be in the research study, you will not be able to participate in this research study.

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

We encourage you to take some time to think this over and to discuss it with other people and to ask questions now and at any time in the future.

B. WHY IS THIS RESEARCH STUDY BEING DONE?

Deaths from cancer in general and for some specific cancers are higher for black people compared to white people, for poor persons compared to nonpoor persons, and for rural residents compared to non-rural residents. There are many reasons for higher death rates between different subpopulations. One important area for changing this is to have more persons from minority groups participate in research about cancer. The process of enrolling minority populations into clinical trials is difficult and does not generally address the needs of their communities. One potential way to increase participation in research is to use community health workers to help educate communities about research and about how to make sure that researchers are ethical. We want to know whether community health workers think this is a good strategy and how to best carry it out.

C. WHAT OTHER OPTIONS ARE THERE?

Taking part in this research study is voluntary. Instead of being in this research study, you have the following option:

- Decide not to participate in this research study.

D. WHAT IS INVOLVED IN THE RESEARCH STUDY?

Before the research starts (screening): After signing this consent form, you will be asked to answer some questions about where you work and the type of community health work you do to find out if you can be in the research study.

If the answers show that you are eligible to participate in the research study, you will be eligible to participate in the research study. If you do not meet the eligibility criteria, you will not be able to participate in this research study.

After the screening procedures confirm that you are eligible to participate in the research study: You will participate in an interview by answering questions from a questionnaire. The interview will take about 90 minutes. If there are questions you prefer not to answer we can skip those questions. The questions are about the type of work you do and your opinions about participating in research. If you agree, the interview will be taped and then transcribed. Your name and no other information about you will be associated with the tape or the transcript. Only the research team will be able to listen to the tapes.

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

Immediately following the interview, you will have the opportunity to have the tape erased if you wish to withdraw your consent to taping or participation in this study. You will receive $30.00 for completing this interview.

After the interview is completed: Once you finish the interview there are no additional interventions.

...

N. DOCUMENTATION OF CONSENT

My signature below indicates my willingness to participate in this research study and my understanding that I can withdraw at any time.

________________________ ________________________

Signature of Subject Date
or Legally Authorized Representative

________________________ ________________________

Person obtaining consent Date

To be completed by person obtaining consent:

The consent discussion was initiated on ________________ (date) at ______________ (time.)

☐ A copy of this signed consent form was given to the subject or legally authorized representative.

For Adult Subjects

☐ The subject is an adult and provided consent to participate.

☐ The subject is an adult who lacks capacity to provide consent and his/her legally authorized representative:

  ☐ gave permission for the adult subject to participate

  ☐ did not give permission for the adult subject to participate

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

Consent Form A was approved by my university IRB for a mailed survey about substance abuse among undergraduate students. It is brief and to the point.

Consent Form B reflects the requirements of an academic hospital's IRB (I have included only a portion of the six-page form). Because the hospital is used to reviewing research proposals involving drugs and other treatment interventions with hospital patients, it requires a very detailed and lengthy explanation of procedures and related issues, even for a simple interview study such as mine with Dr. Schapira. You can probably imagine that the requirement that prospective participants sign such lengthy consent forms can reduce their willingness to participate in research and perhaps influence their responses if they do agree to participate (Larson 1993:114).

A lengthy consent form can also be difficult for participants to understand when it discusses issues like medical treatments and possible side effects. The new federal human subjects regulations adopted in 2017 include a requirement for an initial concise explanation in consent forms that is intended to improve participant understanding (*Federal Register* 2017:7265):

> Informed consent must begin with a concise and focused presentation of the key information that is most likely to assist a prospective subject or legally authorized representative in understanding the reasons why one might or might not want to participate in the research. This part of the informed consent must be organized and presented in a way that facilitates comprehension.

Key information is to include identification of the consent as involving voluntary participation in research, its purposes and expected duration and procedures, and reasonably foreseeable risks and benefits, as well as alternative treatments (Chadwick 2017:5).

As in Milgram's study, experimental researchers whose research design requires some type of subject deception try to get around this problem by withholding some information before the experiment begins but then debriefing subjects at the end. In a debriefing, the researcher explains to the subjects what happened in the experiment and why, and then responds to their questions. A carefully designed debriefing procedure can help the research participants learn from the experimental research and grapple constructively with feelings elicited by the realization that they were deceived (Sieber 1992:39–41). However, even though debriefing can be viewed as a substitute, in some cases, for securing fully informed consent before the experiment, debriefed subjects who disclose the nature of the experiment to other participants can contaminate subsequent results (Adair, Dushenko, and Lindsay 1985). Apparently for this reason, Milgram provided little information in his "debriefing" to participants in most of his experiments. It was only in the last two months of his study that he began to provide more information, while still asking participants not to reveal the true nature of the experimental procedures until after the study was completely over (Perry 2013:76, 84). Unfortunately, if the debriefing process is delayed, the ability to lessen any harm resulting from the deception is also reduced.

Debriefing: A researcher's informing subjects after an experiment about the experiment's purposes and methods and evaluating subjects' personal reactions to the experiment.

For a study of the social background of men who engage in homosexual behavior in public facilities, Laud Humphreys (1970) decided that truly informed consent would be impossible to obtain. Instead, he first served as a lookout—a "watch queen"—for men who were entering a public bathroom in a city park with the intention of having sex. In a number of cases, he then left the bathroom and copied the license plate numbers of the cars driven by the men. One year later, he visited the homes of the men and interviewed them as part of a larger study of social issues. Humphreys changed his appearance so that the men did not recognize him. In *Tearoom Trade,* his book on this research, Humphreys concluded that the men who engaged in what were then viewed as deviant acts were, for the most part, married, suburban men whose families were unaware of their sexual practices. But debate has continued ever since about Humphreys's failure to tell the men what he was really doing in the bathroom or why he had come to their homes for the interview. He was criticized by many, including some faculty members at the University of Washington who urged that his doctoral degree be withheld. However, many other professors and some members of the gay community praised Humphreys for helping normalize conceptions of homosexuality (Miller 1986:135).

Tearoom Trade: The study by sociologist Laud Humphreys of men who engaged in homosexual behavior in public facilities, including subsequent later interviews in their homes after recording their license plate numbers; widely cited in discussions of the need for informed consent to research.

If you were to serve on your university's IRB, would you allow this research to be conducted? Can students who are asked to participate in research by their professor be considered able to give informed consent? Do you consider *informed consent* to be meaningful if the true purpose or nature of an experimental manipulation is not revealed?

The process and even possibility of obtaining informed consent must consider the capacity of prospective participants to give informed consent. Children cannot legally give consent to participate in research; instead, they must in most circumstances be given the opportunity to give or withhold their *assent* to participate in research, usually by a verbal response to an explanation of the research. In addition, a child's legal guardian must give written informed consent to have the child participate in research (Sieber 1992). There are also special protections for other populations who are likely to be vulnerable to coercion or undue influence—defined in the new regulations as prisoners, individuals with impaired decision-making ability, and educationally or economically disadvantaged persons (Chadwick 2017:4; *Federal Register* 2017:7263). Would you allow research on prisoners, whose ability to give informed consent can be questioned? What special protections do you think would be appropriate?

Obtaining informed consent also becomes more challenging in collectivist communities in which leaders or the whole group are accustomed to making decisions for individual members. In such settings, usually in non-Western cultures, researchers may have to develop a relationship with the community before individuals can be engaged in research (Bledsoe and Hopson 2009:397–398). Regulations adopted by the United Nations in 2007 stipulate the right of indigenous communities to control access to research in their communities (Israel 2014:92).

Subject payments create yet another complication for achieving the goal of informed consent. Although payments to research participants can be a reasonable way to compensate them for their time and effort, payments also serve as an inducement to participate. If the payment is a significant amount in relation to the participants' usual income, it could lead people to set aside their reservations about participating in a project—even though they may harbor those reservations (Fisher and Anushko 2008:104–105).

Avoid Deception in Research, Except in Limited Circumstances

Deception occurs when subjects are misled about research procedures to determine how they would react to the treatment if they were not research subjects. Deception is a critical component of many social psychology experiments, partly because of the difficulty of simulating real-world stresses and dilemmas in a laboratory setting. The goal is to get subjects "to accept as true what is false or to give a false impression" (Korn 1997:4). In Milgram's (1963) experiment, for example, deception seemed necessary because the subjects could not be permitted to administer real electric shocks to the "stooge," yet it would not have made sense to order the subjects to do something that they didn't find to be so troubling. Milgram (1992:187–188) insisted that the deception was absolutely essential, although the experimental records indicate that some participants figured out the deception (Perry 2013:128–129). The real question: Is this sufficient justification to allow the use of deception?

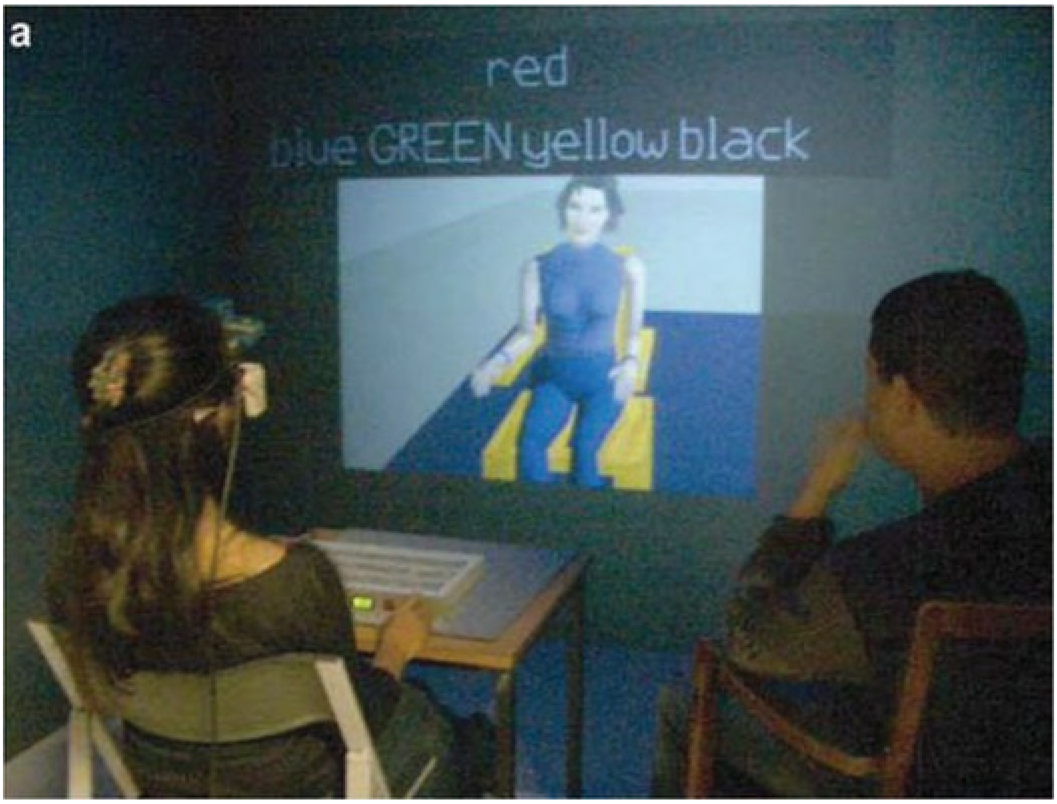
Deception: Used in social experiments to create more "realistic" treatments in which the true purpose of the research is not disclosed to participants, often within the confines of a laboratory.

Gary Marshall and Philip Zimbardo (1979:971–972) sought to determine the physiological basis of emotion by injecting student volunteers with adrenaline, so that their heart rates and sweating would increase, and then placing them in a room with a student "stooge" who acted silly. The students were told that they were being injected with a vitamin supplement to test its effect on visual acuity (Korn 1997:2–3). Jane Allyn Piliavin and Irving Piliavin (1972:355–356) staged fake seizures on subway trains to study helpfulness (Korn 1997:3–4). George Schreer, Saundra Smith, and Kirsten Thomas (2009) investigated racial profiling by sending "customers" to browse in high-end retail stores and then observing the behaviors of salespersons. If you were a member of your university's IRB, would you vote to allow some or all such deceptive practices in research? What about less dramatic instances of deception in laboratory experiments with students like yourself?

The development of computer-based techniques for creating virtual reality environments provides a way to lessen this dilemma. For example, Mel Slater, Angus Antley, and a team of European researchers (2006) repeated the Milgram obedience experiment procedures with virtual reality techniques. According to participants' statements, behaviors, and physiological responses, they seemed to experience the immersive virtual environment as if it was real, even though they clearly understood that the "learner" they observed was only virtual (see Exhibit 3.8). The participants also responded to the "experimenter's" requests in a way that was similar to what occurred in Milgram's "real" experiment.

The new proposed federal regulations relax concerns about deception of the type often used in laboratory experiments in social psychology. Specifically, the regulations authorize deception in research "where the subject is informed that he or she will be unaware of or misled regarding the nature or purposes of the research" (Chadwick 2017:4). Ethics codes in some European countries permit covert research (Israel 2014:99).

Exhibit 3.8 Virtual Reprise of Milgram's Obedience Experiments



Source: Slater M, Antley A, Davison A, Swapp D, Guger C. 2006. “A Virtual Reprise of the Stanley Milgram Obedience Experiments.” PLoS ONE 1(1).

Maintain Privacy and Confidentiality

Maintaining privacy and confidentiality is another key ethical standard for protecting research participants, and the researcher’s commitment to that standard should be included in the informed consent agreement (Sieber 1992). Procedures to protect each subject’s privacy—such as locking records and creating special identifying codes—must be established to minimize the risk of access by unauthorized persons. However, statements about confidentiality should be realistic: Laws allow research records to be subpoenaed and may require reporting of child abuse; a researcher may feel compelled to release information if a health- or life-threatening situation arises and participants need to be alerted. It can be difficult to maintain confidentiality when research is reported about a specific community and involves particular behaviors or events that community members may recognize (Israel 2014:107–113). Also, the standard of confidentiality does not apply to observation in public places and information available in public records.

There is one exception to some of these constraints: The National Institutes of Health (NIH) can issue a Certificate of Confidentiality to protect researchers from being legally required to disclose confidential information. This is intended to help researchers overcome the reluctance of individuals engaged in illegal behavior to sign a consent form or to risk exposure of their illegal activities (Sharma 2009:426). Researchers who are focusing on high-risk populations or behaviors, such as crime, substance abuse, sexual activity, or genetic information, can request such a certificate. Suspicions of child abuse or neglect must still be reported, and in some states, researchers may still be required to report such crimes as elder abuse (Arwood and Panicker 2007).

The Health Insurance Portability and Accountability Act (HIPAA) passed by Congress in 1996 created more stringent regulations for the protection of health care data. As implemented by the DHHS in 2000 (revised in 2002), the HIPAA Final Privacy Rule applies to oral, written, and electronic information that “relates to the past, present or future physical or mental health or condition of an individual.” The HIPAA rule requires that researchers have valid authorization for any use or disclosure of “protected health information” from a health care provider. Waivers of authorization can be granted in special circumstances (Cava, Cushman, and Goodman 2007).

Consider Uses of Research So That Benefits Outweigh Risks

Scientists must also consider the uses to which their research is put. Although many scientists believe that personal values should be left outside the laboratory, some feel that it is proper—even necessary—for scientists to concern themselves with the way their research is used.

Milgram made it clear that he was concerned about the phenomenon of obedience precisely because of its implications for people's welfare. As you have already learned, his first article (Milgram 1963) highlighted the atrocities committed under the Nazis by citizens and soldiers who were "just following orders." In his more comprehensive book on the obedience experiments (Milgram 1974), he also argued that his findings shed light on the atrocities committed in the Vietnam War at My Lai, slavery, the destruction of the American Indian population, and the internment of Japanese Americans during World War II. Milgram makes no explicit attempt to "tell us what to do" about this problem. In fact, as a dispassionate social scientist, Milgram (1974) says, "What the present study [did was] to give the dilemma [of obedience to authority] contemporary form by treating it as subject matter for experimental inquiry, and with the aim of understanding rather than judging it from a moral standpoint" (p. xi).

Certificate of Confidentiality: A certificate issued to a researcher by the National Institutes of Health that ensures the right to protect from legal subpoenas information obtained about high-risk populations or behaviors—except child abuse or neglect.

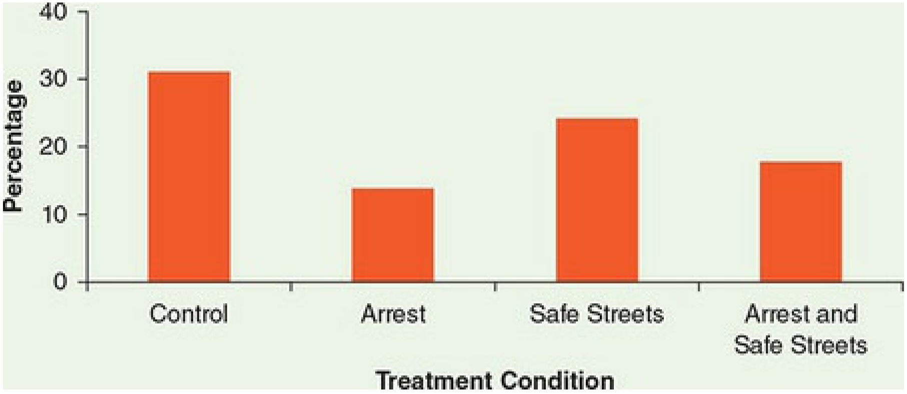
Health Insurance Portability and Accountability Act (HIPAA): Congressional legislation passed in 1996 that creates stringent regulations for the protection of health care data.

Yet it is impossible to ignore the very practical implications of Milgram's investigations, which Milgram took pains to emphasize. His research highlighted the extent of obedience to authority. Although it was widely discussed, Milgram's various versions of the obedience experiment also identified multiple factors that could be manipulated to lessen blind obedience, including encouraging dissent by at least one group member, removing the subject from direct contact with the authority figure, and increasing the contact between the subject and the victim. It is less clear how much Milgram's laboratory manipulation can tell us about obedience in the very different historical events to which he generalized his conclusions, but some subsequent research tried to evaluate this. In the Hofling Hospital Experiment, real nurses obediently administered a lethal dose of a (fake) medicine when requested by a (fake) doctor on the phone (Hofling et al., 1966)—although this was not true when the command was given in person by a real doctor about a real medicine and the nurses had time to consult each other (Rank and Jacobson 1977). Do you agree that this type of research can have potentially great benefits for society?

The evaluation research by Lawrence Sherman and Richard Berk (1984) on police response to domestic violence provides an interesting cautionary tale about the uses of science. As you recall from Chapter 2, the results of this field experiment indicated that those who were arrested were less likely to subsequently commit violent acts against their partners. Sherman (1993) explicitly cautioned police departments not to adopt mandatory arrest policies based solely on the results of the Minneapolis experiment, but the results were publicized in the mass media and encouraged many jurisdictions to change their policies (Binder and Meeker 1993; Lempert 1989). We now know that the original finding of a deterrent effect of arrest did not hold up in many other cities where the experiment was repeated, so it is not clear that the initial changes in arrest policy were beneficial. Sherman (1992:150–153) later suggested that implementing mandatory arrest policies might have prevented some subsequent cases of spouse abuse, but this does not change the fact that these policies were often ineffective.

Given the mixed findings from the replications of Sherman and Berk's experiment, do you think that police policy should be changed in light of JoAnn Miller's (2003) analysis of victims' experiences and perceptions concerning their safety after the mandatory arrest experiment in Dade County, Florida? Miller found that victims reported experiencing less violence after their abuser had been arrested (and/or assigned to a police-based counseling program called Safe Streets) (see Exhibit 3.9). Should this Dade County finding be publicized in the popular press, so that it could be used to improve police policies? What about the results of the other replication studies?

Exhibit 3.9 Victim Reports of Violence Following Police Intervention



Source: Miller, JoAnn. 2003. "An Arresting Experiment: Domestic Violence Victim Experiences and Perceptions." Journal of Interpersonal Violence 18:695–716.

Social scientists who conduct research on behalf of specific organizations may face additional difficulties when the organization, instead of the researcher, controls the final report and the publicity it receives. If organizational leaders decide that particular research results are unwelcome, the researcher's desire to have the findings used appropriately and reported fully can conflict with contractual obligations. Researchers can often anticipate such dilemmas in advance and resolve them when the contract for research is negotiated—or they may simply decline a particular research opportunity altogether. But, often, such problems come up only after a report has been drafted, or the problems are ignored by a researcher who needs to have a job or needs to maintain particular personal relationships. These possibilities cannot be avoided entirely, but because of them, it is always important to acknowledge the source of research funding in reports and to consider carefully the sources of funding for research reports written by others.

The potential of withholding a beneficial treatment from some subjects also is a cause for ethical concern. The Sherman and Berk (1984) experiment required the random assignment of subjects to treatment conditions and thus had the potential of causing harm to the victims of domestic violence whose batterers were not arrested. The justification for the study design, however, is quite persuasive: The researchers didn't know before the experiment which response to a domestic violence complaint would be most likely to deter future incidents (Sherman 1992). The experiment provided what seemed at first to be clear evidence about the value of arrest, so it can be argued that the benefits outweighed the risks.

In some projects the safety of the researchers and their staff also requires consideration. Although it doesn't happen often, research in some social settings about organized crime, and in some countries about insurgent movements and military abuses, have resulted in researchers being killed, injured, or arrested. Interviewers or observers may also suffer from emotional reactions to harm done to others, including those they interview. No research project should begin without careful consideration of whether researcher and staff safety may be an issue. If it is, a safety protocol must be developed in which procedures for ensuring safety are spelled out (Fawcett and Pockett 2015:138–139; Israel 2014:174–179).

The Institutional Review Board

Federal regulations require that every institution that seeks federal funding for biomedical or behavioral research on human subjects have an institutional review board (IRB) that reviews research proposals involving human subjects—including data about living individuals. According to federal regulations [45 CFR 46.102(d)], research is “a systematic investigation . . . designed to develop or contribute to generalizable knowledge,” and according to the DHHS [45 CFR 46.102 (f)], a human subject is “a living individual about whom an investigator (whether professional or student) conducting research obtains data through intervention or interaction with the individual or just identifiable private information.” The IRB determines whether a planned activity is research or involves human subjects.

IRBs at universities and other agencies apply ethical standards that are set by federal regulations but can be expanded or specified by the institution’s IRB and involve all research at the institution irrespective of the funding source (Sieber 1992:5, 10). The IRB has the authority to require changes in a research protocol or to refuse to approve a research protocol if it deems human subjects protections inadequate. Consent forms must include contact information for the IRB, and the IRB has the authority to terminate a research project that violates the procedures the IRB approved or that otherwise creates risks for human subjects. The Office for Protection From Research Risks, National Institutes of Health monitors IRBs, with the exception of research involving drugs (which is the responsibility of the U.S. Food and Drug Administration).

Institutional review board (IRB): A group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to human subjects.

Office for Protection From Research Risks, National Institutes of Health: The office in the U.S. Department of Health and Human Services that provides leadership and supervision about the protection of the rights, welfare, and well-being of subjects involved in research conducted or supported by DHHS, including monitoring IRBs.

To promote adequate review of ethical issues, the regulations require that IRBs include at least five members, with at least one nonscientist and one from outside the institution (Speiglman and Spear 2009:124). The new proposed regulations for IRBs also stipulate that they “shall be sufficiently qualified through the experience and expertise of its members, including race, gender, and cultural backgrounds and sensitivity to such issues as community attitudes, to promote respect for its advice and counsel in safeguarding the rights and welfare of human subjects (*Federal Register* 2017:7263). When research is reviewed concerning populations vulnerable to coercion or undue influence, such as prisoners, the IRB must include a member knowledgeable about that population (Chadwick 2017:4). Sensitivity to community attitudes and training in human subjects protection procedures is also required (Selwitz, Epley, and Erickson 2013).

Every member of an institution with an IRB—including faculty, students, and staff at a college or university—must submit a proposal to their IRB before conducting research with identifiable people. The IRB proposal must include research instruments and consent forms, as applicable, as well as enough detail about the research design to convince the IRB members that the potential benefits of the research outweigh any risks (Speiglman and Spear 2009:124). Most IRBs also require that researchers complete a training program about human subjects, usually the Collaborative Institutional Training Initiative (CITI) at the University of Miami (https://about.citiprogram.org/en/homepage/). CITI training is divided into topical modules ranging from history, ethical principles, and informed consent to vulnerable populations, Internet-based research, educational research, and records-based research. Each IRB determines which CITI training modules researchers at its institution must complete.

In the News Research in the News: Some Social Scientists Are Tired of Asking for Permission

For Further Thought?

The 2017 revision of the 1991 Federal Policy for the Protection of Human Subjects (known as the Common Rule) became quite newsworthy after an opinion piece in *The Chronicle of Higher Education* noted the apparent new exemption from IRB review of research involving "benign behavioral interventions." In the opinion of co-author Richard Nisbett, psychology professor at the University of Michigan, "There's no such thing as asking a question of a normal human being that should be reviewed by an I.R.B., because someone can just say, 'To heck with you.'" In contrast, Tom George, a lawyer and bioethicist on the institutional review board at the University of Texas at Austin worried that "there seems to be a major paradigm shift going on away from . . . protect[ing] human subjects and toward the convenience of researchers." Nathaniel Herr, psychology professor at American University observed, "It just takes one scandal to make people doubt all research and not want to participate, which would harm the whole field."

1. Do you believe that social science researchers should be able to determine whether their research is subject to IRB review?
2. Professor Nisbett felt a "behavioral intervention" is benign "if it's the sort of thing that goes on in everyday life." Do you agree?

News source: Murphy, Kate. 2017. "Some Social Scientists Are Tired of Asking for Permission." *The New York Times,* May 22.

Although the IRB is the responsible authority within the institution (Hicks 2013), in the new proposed regulations many proposals developed by social scientists will be exempt from review because they involve very low perceived risk. These exemptions include (1) research about educational procedures in an educational setting, or (2) "educational tests, survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording)," or (3) "research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses . . . or audiovisual recording" when the subject had agreed to the intervention (*Federal Register* 2017:7261, 7262, 7264). Criteria (2) and (3) also include the stipulation that at least one of the following criteria is met:

- The identity of the human subjects *cannot* readily be ascertained.
- Disclosure of the human subjects' responses would not place them at risk of criminal or civil liability or damage their financial standing, employability, educational advancement, or reputation.
- The identity of the human subjects *can* readily be ascertained, but a limited IRB review determines that procedures are in place that will protect the privacy of subjects and maintain the confidentiality of data.

Other projects must be reviewed before the full IRB (Speiglman and Spear 2009:125–126). An IRB must ensure that several specific standards are met by research projects that it reviews either on an expedited basis or in a full board review (Hicks 2013):

1. Risks to subjects are minimized: (i) By using procedures that are consistent with sound research design and that do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.
2. Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility.
3. Selection of subjects is equitable. In making this assessment the IRB should consider the purposes of the research and the setting in which the research will be conducted and should be particularly cognizant of the special problems of research involving vulnerable populations.

Informed consent is also required and, when appropriate, a provision for monitoring data collection to ensure subject safety. In addition, when some or all of the subjects are likely to be vulnerable to coercion or undue influence, additional safeguards have been included in the study to protect the rights and welfare of these subjects.

In addition, the IRB may serve as the privacy board that ensures researchers' compliance with HIPAA. In this capacity, the IRB responds to requests for waivers or alterations of the authorization requirement under the privacy rule for uses and disclosures of protected health information in research. Researchers seeking to collect or use existing HIPAA data must provide additional information to the IRB about their plans for using the health information.

Careers and Research

Manan Nayak, Senior Project Director

After Manan Nayak graduated from the accelerated BA/MA program in applied sociology at the University of Massachusetts Boston, she began her career as a quality assurance analyst for a university-affiliated medical center. Initially, she used her quantitative skills to manage data from multiple clinical trials. In this role, she submitted regular reports to various committees, including the data safety and monitoring committee that ensures each study is scientific and ethically appropriate based on federal regulations. However, it was not until she became a clinical researcher that she appreciated the importance of human subjects boards. As she approached eligible patients for studies, she learned that many patients wanted to participate in the hopes that the data collected could help someone else—despite already dealing with the effects of treatment and multiple demands on their time. The patients' selflessness motivated Nayak to develop her research career and learn more about ethical and regulatory issues and how to ensure that research teams adhere to strict guidelines. She worked alongside investigators to write applications that clearly state the process the research team will follow, including how participants are identified, what they will be asked to consent to and for how long, as well as how their data will be collected, stored, and distributed. In her work, she ensures that the procedures outlined and approved by the regulatory boards are followed strictly, and any major or minor deviations are reported to the IRB immediately, along with a resolution indicating how infractions can be avoided in the future. Bringing to fruition a research study and making a small contribution in understanding how a treatment affects a group of patients and the challenges they face during treatment are the rewards of doing such research. Nayak's advice to future researchers is to recognize the excitement of doing social research and the many opportunities available to apply skills you learn in research courses.

Social Research Proposals

Now that you have an overview of the research process and a basic understanding of IRB requirements, it is time to introduce the process of writing a research proposal. A research proposal is the launching pad for a formal research project, and it serves the very important function of forcing you, as the researcher, to set out a problem statement and a research plan—to think through the details of what you are trying to accomplish and how you will go about that—as well as to think through procedures for the protection of human subjects. So whether you must write a proposal for a professor, a thesis committee, an organization seeking practical advice, or a government agency that funds basic research, you should approach the requirement as a key step toward achieving your goals. Just writing down your ideas will help you see how they can be improved, and almost any feedback will help you refine your plans.

Each chapter in this book includes a section, "Developing a Research Proposal," with exercises that guide you through the process of proposal writing. This section introduces the process of proposal writing. It also provides a schematic overview of the entire research process. You will want to return to this section frequently so that you will remember "where you are" in the research process as you learn about particular methods in the remaining chapters.

Research proposals often have five sections (Locke, Spirduso, and Silverman 2000:8–34):

- *An introductory statement of the research problem,* in which you clarify what it is that you are interested in studying
- *A literature review,* in which you explain how your problems and plans build on what has already been reported in the literature on this topic
- *A methodological plan,* detailing just how you will respond to the particular mix of opportunities and constraints you face
- *An ethics statement,* identifying human subjects issues in the research and how you will respond to them in an ethical fashion
- *A statement of limitations,* reviewing the potential weaknesses of the proposed research design and presenting plans for minimizing their consequences

You will also need to include a budget and project timeline, unless you are working within the framework of a class project.

When you develop a research proposal, it will help to ask yourself a series of questions such as those in Exhibit 3.10; see also Gregory Herek (1995). It is easy to omit important details and to avoid being self-critical while rushing to put together a proposal. However, it is even more painful to have a proposal rejected (or to receive a low grade). It is better to make sure the proposal covers what it should and confronts the tough issues that reviewers (or your professor) will be sure to consider.

The series of questions in Exhibit 3.10 can serve as a map to subsequent chapters in this book and as a checklist of decisions that must be made throughout any research project. The questions are organized in five sections, each concluding with a checkpoint at which you should consider whether to proceed with the research as planned, modify the plans, or stop the project altogether. The sequential ordering of these questions obscures a bit the way in which they should be answered: not as single questions, one at a time, but as a unit —first as five separate stages and then as a whole. Feel free to change your answers to earlier questions on the basis of your answers to later questions.

We will learn how to apply the decision checklist with an example from a proposal focused on a public health care coordination program. At this early point in your study of research methods, you may not recognize all the terms in this checklist. Don’t let that bother you now because my goal is just to give you a quick overview of the decision-making process. Your knowledge of these terms and your understanding of the decisions will increase as you complete each chapter. Your decision-making skills will also improve if you complete the “Developing a Research Proposal” exercises at the end of each chapter.

Exhibit 3.10 Decisions in Research

Problem Formulation (Chapters 1–3)

1. Developing a research question
2. Assessing researchability of the problem
3. Consulting prior research
4. Relating to social theory
5. Choosing an approach: Deductive? Inductive? Descriptive?
6. Reviewing research guidelines and ethical standards

Checkpoint 1
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Research Validity (Chapters 4–6)

7. Establishing measurement validity:
    - How should concepts be defined?
    - What measures are available, and what measures must be developed?
    - How can reliability and validity be assessed?
    - Is authenticity an important goal, and how can it be assessed?
8. Establishing generalizability:
    - Was a representative sample used?
    - Are the findings applicable to particular subgroups?
    - Does the population sampled correspond to the population of interest?
9. Establishing causality:
    - What is the possibility of experimental or statistical controls?
    - How can the causal mechanism be assessed?
    - What is the causal context?
10. Determining the data required: Longitudinal or cross-sectional?
11. Determining the units of analysis: Individuals or groups?
12. Determining the major possible sources of causal invalidity

Checkpoint 2
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Research Design (Chapters 7, 8, 10, 12–15)

13. Choosing a research design and procedures: Experimental? Survey? Participant observation? Historical, comparative? Evaluation research? Secondary data analysis? Mixed methods?
14. Specifying the research plan: Type of surveys, observations, etc.
15. Determining secondary analysis and availability of suitable data sets
16. Choosing a causal approach: Idiographic or nomothetic?
17. Assessing human subjects protections

Checkpoint 3
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Data Analysis (Chapters 9, 11)

18. Choosing an analytic approach:
    - Identifying statistics and graphs for describing data
    - Identifying relationships between variables
    - Deciding about statistical controls
    - Testing for interaction effects
    - Evaluating inferences from sample data to the population
    - Developing a qualitative analysis approach

Checkpoint 4
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Reporting Research (Chapter 16)

19. Clarifying research goals and prior research findings
20. Identifying the intended audience
21. Developing tables and charts
22. Organizing the text
23. Reviewing research limitations

Checkpoint 5
Alternatives:

- Complete the research!
- Modify the plan.
- Stop. Abandon the plan.

Case Study: Evaluating a Public Health Program

Exhibit 3.11 provides excerpts from the research proposal I submitted to our IRB as part of an evaluation of a public health program for low-income residents funded by the U.S. Centers for Disease Control and the Massachusetts Department of Public Health (DPH) (Schutt 2011a). Appendixes included consent forms, research instruments, and the bibliography.

As you can see from the excerpts, I proposed to evaluate a care coordination program for low-income uninsured and underinsured Massachusetts residents (before universal health care). The proposal included a lengthy literature review, a description of the population and the sampling procedure, measures to be used in the survey, and the methods for conducting phone interviews as well as in-person interviews with a subset of the sample. Required sections for the IRB also included a statement of risks and benefits, procedures for obtaining informed consent, and means for maintaining confidentiality. A HIPAA compliance statement was included because of the collection of health-related data.

Exhibit 3.11 An IRB Proposal for a Program Evaluation

Evaluation of the Coordinated Care Program

Rationale

Inadequate personal resources, knowledge, and service opportunities limit the access of low-income and minority populations to health care services. Percac-Lima and colleagues (2010) identified four types of barriers to effective service delivery in interviews with Latina women using a health care center: anxiety or fear and concerns about pain associated with the procedure ("personal barriers") and inadequate communication and scheduling difficulties ("system barriers"). Vargas et al. (2008) identified additional systemic barriers of provider bias and poor quality of care. Many public health agencies have adopted programs designed to overcome these barriers and thus to reduce disparities in health care services and, ultimately, outcomes.

The Massachusetts Department of Public Health began a Coordinated Care Program in 2006 in order to improve identification of health needs among low-income uninsured and underinsured residents and to increase the effectiveness of health services for this population. . . .

The purpose . . . is to investigate patient experience with the program, including subjective outcomes of the program and the bases of these outcomes. . . .

Program Background

Congress funded the National Breast and Cervical Cancer Early Detection Program (Breast and Cervical Cancer Mortality Prevention Act of 1990) to improve the rate of screening, testing, and referral to treatment for low-income uninsured and underinsured women at risk of breast and cervical cancer . . .

. . . . Programs offering patient navigation and care coordination have not typically been combined together due to their different disease foci: cancer for patient navigation and chronic illness for care coordination. The MA CCP program thus provides a unique opportunity to study the effect of combining these two different types of programs and examining their value for a very diverse and historically underserved population.

HIPAA Compliance Information

. . . No information about health conditions will be reported in a way that could be linked to any particular client.

Methodology

. . . Data will be collected with a statewide phone survey of a representative sample and through in-person interviews with a small number of clients at six of the 17 program delivery sites.

The phone survey is to result in a total sample of 400 interviews. One thousand clients will be sent an informational letter about the survey, indicating that they may receive a call. Cases will be selected randomly from the initial sampling frame and called to determine study eligibility. Phone calls will be repeated up to 20 times in order to make contact. Questions S0-S4 (in the attached questionnaire) will serve as the screening questions. . . . The phone survey instrument will be translated into Spanish and Portuguese. . . .

Questions in the phone survey instrument are designed to identify relation to the health center and patient navigator, . . . perceived barriers to the use of health services, and satisfaction with health services. . . .

A sample of 30 clients will be selected for more intensive interviews about program experience and orientations. . . . For those clients who consent, the in-person interviews will be recorded and transcribed. The resulting textual data will be coded and analyzed with the assistance of qualitative analysis software . . .

Human Participants Information

Participant Data

The client database maintained by the Massachusetts Care Coordination Program identifies . . . active clients. . . .

Risks and Benefits

Participation in the Client Survey involves a minimal risk of loss of confidentiality. In addition, the probability and magnitude of harm or discomfort anticipated from responding to the Client Survey items is no greater than what would ordinarily be encountered in daily life, that is, minimal risk. Although the only direct benefit to survey participants is the opportunity to express their perceptions of and satisfaction with . . . case management, the information gained may be helpful to case managers and other health care providers. . . . The potential benefits of the proposed survey outweigh the minimal risk for individual survey participants.

Informed Consent and Confidentiality

The . . . phone survey interviewer will inform the clients that the survey is voluntary and explain procedures to be followed (the respondent can stop at any time; can skip any question she does not want to answer). . . . . [P]hone interviewers and other staff adhere to strict confidentiality requirements. Identifying information will be used only for client contact and no personal identifiers will be maintained in the phone survey database. The phone survey ID numbers will be stored in a locked file . . . All in-person interviews will be taped and transcribed and records will also be stored in a locked cabinet and on password-protected computers. . . . Consent forms will include consent for the interview, consent for taping, and consent for follow-up.

Let's review the issues identified in Exhibit 3.10 as they relate to the public health proposal. The research question concerned the effectiveness of a care coordination program involving the use of patient navigators—an evaluation research question [Question 1]. This problem certainly was suitable for social research, and it was one that was feasible with the money DPH had committed [2]. Prior research demonstrated clearly that the program had potential but also that this approach had not previously been studied [3]. The treatment approach was connected to theories about health care disparities [4] and, given prior work and uncertainties in this area, mixed methods involving both a deductive, hypothesis-testing approach and an inductive, exploratory approach were called for [5]. I argued to the IRB that our plan protected human subjects and took each research guideline into account [6]. So it seemed reasonable to continue to develop the proposal (Checkpoint 1).

Measures were to include structured survey questions to measure many variables and open-ended questions to allow exploration of barriers to care [7]. Use of a representative sample of the population of program recipients would increase the generalizability of the findings, although I was limited to interviews in English, Spanish, and Portuguese even though we knew there were a small number of recipients who spoke other languages [8]. The problem was well suited to a survey design in which questions about causality could be addressed with statistical controls, with some attention to contextual differences between different service sites. However, there were to be no tests of causal hypotheses or investigation of causal mechanisms, since the program was already in place and there was no opportunity to design a randomized experiment of the effect of the new program [9]. Cross-sectional data [10], involving individuals [11], could be used to investigate reactions to the program. The design left several sources of causal invalidity, including the possibility that persons who received the most services from patient navigators were those who had more resources and so were more healthy [12]. It seemed that I would be able to meet basic criteria for validity, but there would be uncertainty about the causal effect of the new program features (Checkpoint 2).

A survey design was preferable because this was to be a study of a statewide population, but I did include a qualitative component so that I could explore orientations [13, 14]. Because the effectiveness of the program strategy had not been studied before in this type of population, I could not propose doing a secondary data analysis or meta-analysis [15]. I sought only to investigate causation from a *nomothetic* perspective, without attempting to show how the particular experiences of each participant may have led to their outcome [16]. The study design was low risk and included voluntary participation; the research design seemed ethical [17] (Checkpoint 3). Standard statistical tests were proposed as well as some analysis of qualitative data [18] (Checkpoint 4). My goal was to use the research as the basis for several academic articles, as well as a report to the agency [19, 20]. Of course at this point I was not ready to develop tables and charts [21] or organize the text of the report [22], but I was starting to think about study methodology's limitations [23] (Checkpoint 5).

If your research proposal will be reviewed competitively, it must present a compelling rationale for funding. You should not overstate the importance of the research problem that you propose to study (see the first section of this chapter). If you propose to test a hypothesis, be sure that it is one for which there are plausible alternatives. You want to avoid focusing on a "boring hypothesis"—one that has no credible alternatives, even though it is likely to be correct (Dawes 1995:93).

A research proposal also can be strengthened considerably by presenting results from a pilot study of the research question. This might have involved administering the proposed questionnaire to a small sample, conducting a preliminary version of the proposed experiment with a group of students, or making observations over a limited period in a setting like that proposed for a qualitative study. My original proposal to the DPH was strengthened by my ability to build on a prior study I had conducted in the agency 10 years previously. Careful presentation of the methods used in the pilot study and the problems that were encountered will impress anyone who reviews the proposal.

Don't neglect the procedures for the protection of human subjects. Even before you begin to develop your proposal, you should find out what procedure your university's IRB requires for the review of student research proposals. Follow these procedures carefully, even if they require that you submit your proposal for an IRB review. No matter what your university's specific requirements are, if your research involves human subjects, you will need to include in your proposal a detailed statement that describes how you will adhere to these requirements.

By the book's end, in Chapter 16, you will have attained a much firmer grasp of the various research decisions outlined in Exhibit 3.10.

Conclusions

The extent to which ethical issues are a problem for researchers and their subjects varies dramatically with the type of research design. Survey research, in particular, creates few ethical problems. In fact, researchers from Michigan's Institute for Survey Research interviewed a representative national sample of adults some years ago and found that 68% of those who had participated in a survey were somewhat or very interested in participating in another; the more times respondents had been interviewed, the more willing they were to participate again. Presumably, they would have felt differently if they had been treated unethically (Reynolds 1979:56–57). Conversely, some experimental studies in the social sciences that have put people in uncomfortable or embarrassing situations have generated vociferous complaints and years of debate about ethics (Reynolds 1979; Sjoberg 1967).

The evaluation of ethical issues in a research project should be based on a realistic assessment of the overall potential for harm and benefit to research subjects rather than an apparent inconsistency between any particular aspect of a research plan and a specific ethical guideline. For example, full disclosure of "what is really going on" in an experimental study is unnecessary if subjects are unlikely to be harmed. Nevertheless, researchers should make every effort to foresee all possible risks and to weigh the possible benefits of the research against these risks. Researchers should consult with individuals with different perspectives to develop a realistic risk–benefit assessment and should try to maximize the benefits to, as well as minimize the risks for, subjects of the research (Sieber 1992:75–108).

Ultimately, these decisions about ethical procedures are not just up to you, as a researcher, to make. Your university's IRB sets the human subjects protection standards for your institution and may require that you submit any research proposal to the IRB for review. So, I leave you with the instruction to review the human subjects guidelines of the ASA or other professional associations in your field, consult your university's procedures for the conduct of research with human subjects, make human subjects protections an important part of your research proposal, and then proceed according to the appropriate standards.

You should now also understand how important it is to plan all aspects of your research systematically and to present this plan in a formal research proposal. Human subjects protections are a vital part of any research plan that includes collection of data from living people, but writing a research proposal is itself a way to protect yourself from the many problems that will arise if you do not think through carefully each stage of the research process before you start to collect data.

You are ready to begin planning the particulars of a research project.

SAGE edge™

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Belmont Report 71
Beneficence 71
Certificate of Confidentiality 85
Code of Ethics (American Sociological Association) 72
Conflict of interest 75
Debriefing 82
Deception 83
Federal Policy for the Protection of Human Subjects 72
Health Insurance Portability and Accountability Act (HIPAA) 85
Institutional review board (IRB) 87
Justice 71
Milgram's obedience experiments 70
Nuremberg War Crime Trials 70
Office for Protection From Research Risks, National Institutes of Health 87
Respect for persons 71
Tearoom Trade 82
Tuskegee Study of Untreated Syphilis in the Negro Male 70
Zimbardo prison simulation study 77

Highlights

- Stanley Milgram's obedience experiments led to intensive debate about the extent to which deception could be tolerated in social science research and how harm to subjects should be evaluated.
- Egregious violations of human rights by researchers, including scientists in Nazi Germany and researchers in the Tuskegee syphilis study, led to the adoption of federal ethical standards for research on human subjects.
- The 1979 *Belmont Report* developed by a national commission established three basic ethical standards for the protection of human subjects: (1) respect for persons, (2) beneficence, and (3) justice.
- The U.S. Department of Health and Human Services adopted in 1991 the Federal Policy for the Protection of Human Subjects. This policy requires that every institution seeking federal funding for biomedical or behavioral research on human subjects have an institutional review board (IRB) to exercise oversight.
- Current standards for the protection of human subjects require avoiding harm, obtaining informed consent, avoiding deception except in limited circumstances, maintaining privacy and confidentiality, and ensuring that the benefits of research outweigh foreseeable risks.
- The American Sociological Association's general principles for professional practice urge sociologists to be committed in their work to high levels of competence, to practicing with integrity, and to maintaining responsibility for their actions. They must also respect the rights, dignity, and diversity of others, including research participants, as well as be socially responsible to their communities and use research to contribute to the public good.
- Scientific research should maintain high standards for validity and be conducted and reported in an honest and open fashion.

- Effective debriefing of subjects after an experiment can help reduce the risk of harm resulting from the use of deception in the experiment.
- Writing a research proposal is an important part of preparing for research. Key decisions can be viewed as checkpoints that will shape subsequent stages. Proposals may need to be submitted to the university IRB for review before the research begins.

Discussion Questions

1. Should social scientists be permitted to conduct replications of Milgram’s obedience experiments? Zimbardo’s prison simulation? Can you justify such research as permissible within the current ASA ethical standards? If not, do you believe that these standards should be altered to permit Milgram-type research?
2. How do you evaluate the current ASA ethical code? Is it too strict or too lenient, or just about right? Are the enforcement provisions adequate? What provisions could be strengthened?
3. Why does unethical research occur? Is it inherent in science? Does it reflect “human nature”? What makes ethical research more or less likely?
4. Does debriefing solve the problem of subject deception? How much must researchers reveal after the experiment is over as well as before it begins?
5. What policy would you recommend that researchers such as Sherman and Berk (1984) follow in reporting the results of their research? Should social scientists try to correct misinformation in the popular press about their research, or should they just focus on what is published in academic journals? Should researchers speak to audiences such as police conventions to influence policies related to their research results?

Practice Exercises

1. Pair up with one other student and read the article by John Lacey and others, or another from the Research That Matters vignettes in the preceding two chapters. One of you should criticize the research in terms of its adherence to each of the ethical principles for research on human subjects, as well as for the authors' apparent honesty, openness, and consideration of social consequences. Be generally negative but not unreasonable in your criticisms. The other one of you should critique the article in the same way but from a generally positive standpoint, defending its adherence to the five guidelines, but without ignoring the study's weak points. Together, write a summary of the study's strong and weak points, or conduct a debate in class.
2. Investigate the standards and operations of your university's IRB. Review the IRB website, record the composition of the IRB (if indicated), and outline the steps that faculty and students must take to secure IRB approval for human subjects research. In your own words, distinguish the types of research that can be exempted from review, that qualify for expedited review, and that require review by the full board. If possible, identify another student or a faculty member who has had a proposal reviewed by the IRB. Ask him or her to describe the experience and how he or she feels about it. Would you recommend any changes in IRB procedures?
3. Choose one of the four "Ethical Issues" lessons from the opening menu for the interactive exercises. Review issues in ethical practice by reading the vignettes and entering your answers when requested. You have two chances to answer each question.
4. Also from the book's study site, at edge.sagepub.com/schutt9e, choose the "Learning From Journal Articles" option. Read one article based on research involving human subjects. What ethical issues did the research pose, and how were they resolved? Does it seem that subjects were appropriately protected?

Ethics Questions

1. Lacey and his collaborators in the National Roadside Survey (2011), described in the Research That Matters vignette, conducted a purely descriptive study of the prevalence of impaired driving. What if they had sought to test the impact of conducting such traffic stops on the subsequent likelihood of drivers drinking (or drugging) and driving? If this had been their goal, they might have proposed conducting the traffic stops at randomly determined locations, while conducting surveys without a test for impaired driving at other randomly determined locations as their control condition. They could then have followed up a year later to see if those in the traffic stop group were less likely to have been arrested for DUI. The results of such a study could help devise more effective policies for reducing driving under the influence. Do you think an IRB should approve a study like this with a randomized design? Why or why not?
2. Milgram's research on obedience to authority has been used to explain the behavior of soldiers charged with intentionally harming civilians during armed conflicts, both on the battlefield and when guarding prisoners of war. Do you think social scientists can use experiments such as Milgram's to learn about ethical behavior in the social world in general? What if it seems that the same issue can be studied in simulations using virtual reality technology?

Web Exercises

1. The U.S. Department of Health and Human Services maintains extensive resources concerning the protection of human subjects in research. Read several documents that you find on the website of the Office for Human Research Protections, www.hhs.gov/ohrp, and write a short report about them.
2. Read the entire ASA *Code of Ethics* at the website of the ASA Ethics Office, www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf. Discuss the difference between the aspirational goals and the enforceable rules.

Video Interview Questions

Listen to the researcher interview for Chapter 3 at edge.sagepub.com/schutt9e.

1. What are the key issues that an institutional review board (IRB) evaluates in a research proposal?
2. What are some challenges that an IRB faces? How does Dr. Nestor suggest that these challenges can be resolved?

SPSS Exercises

1. Consider three variables in the GSS2016 survey: helpblk, income06, and owngun. Review the corresponding variable labels in the “Variable View” in the GSS2016 file you are using so that you know what questions were used to measure these variables. Are there any ethical issues involved in asking these questions? Do you imagine that some respondents would be more likely to give untruthful answers to these questions? Explain your answers.
2. Examine the cross-tabulation between fundamentalist beliefs and support for the right to abortion after a rape and support for allowing an antireligionist to speak (Analyze→Descriptive Statistics→Crosstabs→ (Rows=abrape, spkath; Columns=fund; Cells/Percentages=columns). Summarize what you have learned about the association between fundamentalist beliefs and feelings about abortion.

Developing a Research Proposal

The following steps add to the very critical first steps identified in Exhibit 3.10 that you have already completed in the research proposal exercises for Chapters 1 and 2:

1. Identify each of the elements of your planned research that might be of concern to an IRB. These could include procedures for drawing a sample, inclusion of particular questions in a survey, or openness about procedures in an experiment.
2. Write an annotated list for the application to the IRB in which you explain how you will ensure that your research adheres to each relevant ASA standard.
3. Draft a consent form to be administered to your research participants when they enroll in your research. Use underlining and marginal notes to indicate where each standard for informed consent statements is met.

Chapter 4 Conceptualization and Measurement

- Research That Matters, Questions That Count
- Concepts
  - Conceptualization in Practice
    - *Substance Abuse*
    - *Youth Gangs*
    - *Poverty*
- From Concepts to Indicators
- Research in the News: Are Teenagers Replacing Drugs With Smartphones?
  - Abstract and Concrete Concepts
    - *Operationalizing the Concept of Race*
    - *Operationalizing Social Network Position*
- From Observations to Concepts
- Measurement
  - Constructing Questions
  - Making Observations
  - Collecting Unobtrusive Measures
  - Using Available Data
  - Coding Content
  - Taking Pictures
  - Combining Measurement Operations
- Careers and Research
- Levels of Measurement
  - Nominal Level of Measurement
  - Ordinal Level of Measurement
  - Interval Level of Measurement
  - Ratio Level of Measurement
  - The Special Case of Dichotomies
  - Comparison of Levels of Measurement
- Evaluating Measures
  - Measurement Validity
    - *Face Validity*
    - *Content Validity*
    - *Criterion Validity*
    - *Construct Validity*
  - Measurement Reliability
    - *Multiple Times: Test–Retest and Alternate Forms*
    - *Multiple Indicators: Interitem and Split-Half*
    - *Multiple Observers: Interobserver and Intercoder*

Ways to Improve Reliability and Validity
Conclusions

Research That Matters, Questions That Count

Excessive use of alcohol, illicit drugs, and cigarettes are not only associated with problems at the time of intoxication, but they also predict long-term differences in the life course. Bohyun Joy Jang and Megan Patrick at the University of Michigan and Megan Schuler at the Harvard Medical School focused attention on the consequences of substance use among young adults for the timing of their subsequent marriage or cohabitation and parenthood.

The primary research question was whether substance use by young adults predicts delays in family formation. Secondarily, the researchers were interested in variation in these patterns between substances used and between males and females, as well as what might reflect preexisting differences between the young adults who start using substances and those who do not. They investigated these research questions with data collected from young adults over a 10-year period in the Monitoring the Future survey.

The concept of substance use exposure was measured at each time point with three questions about frequency of smoking cigarettes, binge drinking, and marijuana use. Measures of the concept of family formation outcomes were questions about marital, cohabitation, and parental status. The researchers also took into account in their analysis responses to questions about race/ethnicity, religiosity, parental education, high school grade point average, and college plans.

You can read more about the measures used and the researchers' conclusions in, respectively, the "Measures" and "Discussion" sections of the Jang et al. article, at the *Investigating the Social World* study site. I will refer to their measures throughout this chapter.

1. What is *your* concept of "risky substance use"? State a tentative definition.
2. How well do you think substance use can be measured with one question each about cigarette use, binge drinking, and marijuana use?
3. Jang et al. focused on young adults aged 21–30 years old. Do you agree with their concern that they could have missed important differences in family formation among youth under the age of 21 and for adults older than 30? Does your own "family formation experience" or that of your family and friends give you any ideas about this?

In this chapter, you will learn about concepts and measures used in studies of substance abuse, youth gangs, poverty, and other social phenomena. By the end of the chapter, you will understand why defining concepts and developing measures are critical steps in research and you will have a much firmer basis for answering the questions I have posed. As you read the chapter, review the details about measurement in Jang et al.'s study of substance use and family formation at the *Investigating the Social World* study site and then test yourself by completing the related interactive exercises for Chapter 4 at edge.sagepub.com/schutt9e. You can also learn more about the Monitoring the Future study at http://www.monitoringthefuture.org/.

Jang, Bohyun Joy, Megan E. Patrick, and Megan S. Schuler. 2017. "Substance Use Behaviors and the Timing of Family Formation During Young Adulthood." *Journal of Family Issues,* Online first.

Substance abuse is a social problem of remarkable proportions, both on and off campus. About 18 million Americans have an alcohol use disorder (Grant et al. 2004; Hasin et al. 2007; National Institute on Alcohol Abuse and Alcoholism [NIAAA] 2013). In the United States, about 80,000 people die every year from alcohol-related causes (NIAAA 2013), as do about 2.5 million annually around the globe (World Health Organization [WHO] 2013). Almost one quarter of young people aged 19–20 in the United States report "binge drinking" (5+ drinks on one occasion) and 1 in 10 report "high-intensity drinking" (10+ drinks on one occasion); as you have just learned from the research by Jang et al. (2017), this can in turn have a long-term influence on family formation. The problem is worse among four-year college students: 29.3% binge drink as compared to 21.5% of those not attending college (Patrick and Terry-McElrath 2017). Drinking is a factor in at least half of on-campus sexual assaults (Abbey 2002). All told, the annual costs of prevention and treatment for alcohol and drug abuse exceed $340 billion in the United States (Miller and Hendrie 2008).

Whether your goal is to learn how society works, to deliver useful services, to design effective social policies, or simply to try to protect yourself and your peers, at some point you might decide to read some of the research literature on substance abuse. Perhaps you will even attempt to design your own study of it. Every time you begin to review or design relevant research, you will have to answer two questions: (1) What is meant by *substance abuse* in this research? (conceptualization) and (2) How was substance abuse measured? (operationalization). Both types of questions must be answered when we evaluate prior research, and both types of questions must be kept in the forefront when we design new research. Only when we conclude that a study used valid measures of its key concepts can we have some hope that its conclusions are valid.

In this chapter, I first address the issue of conceptualization, using substance abuse and other concepts as examples. I then focus on measurement, reviewing first how indicators of substance abuse and several other concepts have been constructed using such operations as questions, observations, and less direct and obtrusive measures. Next I discuss the different possible levels of measurement and methods for assessing the validity and reliability of measures. The final topic to consider is the unique insights that qualitative methods can add to the measurement process. By the chapter's end, you should have a good understanding of measurement, the first of the three legs on which a research project's validity rests.

Concepts

Although the drinking statistics sound scary, we need to be clear about what they mean before we march off to a Temperance Society meeting. What, after all, is binge drinking? The definition that Henry Wechsler et al. (2002) used is “heavy episodic drinking”; more specifically, “we defined binge drinking as the consumption of at least 5 drinks in a row for men or 4 drinks in a row for women during the 2 weeks before completion of the questionnaire” (p. 205).

Is this what you call *binge drinking?* This definition is widely accepted among social researchers, so when they use the term they can understand each other. However, the NIAAA (College Alcohol Study 2008) provides a more precise definition: “A pattern of drinking alcohol that brings blood alcohol concentration to 0.08 grams percent or above.” Most researchers consider the so-called 5/4 definition (5 drinks for men; 4 for women) to be a reasonable approximation of this more precise definition. We can’t say that only one definition of *binge drinking* is “correct,” or even that one is “better.” What we can say is that we need to specify what we mean when we use the term. We also have to be sure that others know what definition we are using. And of course, the definition has to be useful for our purposes: A definition based solely on blood alcohol concentration will not be useful if we are not taking blood measures.

We call binge drinking a concept—a mental image that summarizes a set of similar observations, feelings, or ideas. To make that concept useful in research (and even in ordinary discourse), we have to define it. Many concepts are used in everyday discourse without consistent definition, sometimes definitions of concepts are themselves the object of intense debate, and the meanings of concepts may change over time. For example, when we read a *New York Times* article (Stille 2000) announcing a rise in the “social health” of the United States, after a precipitous decline in the 1970s and 1980s, we don’t know whether we should feel relieved or disinterested. In fact, the authorities on the subject didn’t even agree about what the term *social health* meant: lessening of social and economic inequalities (Marc Miringoff) or clear moral values (William J. Bennett). Most agreed that social health has to do with “things that are not measured in the gross national product” and that it is “a more subtle and more meaningful way of measuring what’s important to [people]” (Stille 2000:A19), but the sparks flew over whose conceptualization of social health would prevail.

Concept: A mental image that summarizes a set of similar observations, feelings, or ideas.

Prejudice is an interesting example of a concept whose meaning has changed over time. As

Harvard psychologist Gordon Allport (1954) pointed out, during the 1950s many people conceptualized prejudice as referring to "faulty generalizations" about other groups. The idea was that these cognitive "errors in reasoning" could be improved with better education. But by the end of the 1960s, this one-size-fits-all concept was replaced with more specific terms such as *racism, sexism,* and *anti-Semitism* that were conceptualized as referring to negative dispositions about specific groups that "ran too deep to be accessible to cursory introspection" (Nunberg 2002:WK3). The *isms* were conceived as both more serious and less easily acknowledged than prejudice.

Concepts such as social health, prejudice, and even binge drinking require an explicit definition before they are used in research because we cannot be certain that all readers will share a particular definition or that the current meaning of the concept is the same as it was when previous research was published. It is especially important to define clearly any concepts that are abstract or unfamiliar. When we refer to concepts such as *social control, anomie,* or *social health,* we cannot count on others knowing exactly what we mean. Even experts may disagree about the meaning of frequently used concepts if they have based their conceptualizations on different theories. That's okay. The point is not that there can only be one definition of a concept but that we have to specify clearly what we mean when we use a concept, and we must expect others to do the same.

Conceptualization in Practice

If we are to do an adequate job of conceptualizing, we must do more than just think up some definition, any definition, for our concepts (Goertz 2006). We have to turn to social theory and prior research to review appropriate definitions. We need to identify what we think is important about the phenomenon that interests us. We should understand how the definition we choose fits within the theoretical framework guiding the research, and what assumptions underlie this framework. We may decide the concept has several dimensions, or subconcepts, that should be distinguished.

Substance Abuse

What observations or images should we associate with the concept *substance abuse?* Someone leaning against a building with a liquor bottle, barely able to speak coherently? College students drinking heavily at a party? Someone in an Alcoholics Anonymous group drinking one beer? A 10-year-old boy drinking a small glass of wine in an alley? A 10-year-old boy drinking a small glass of wine at the dinner table in France? Do all these images share something in common that we should define as substance abuse for the purposes of a particular research study? Do only some of them share something in common? Should we consider the cultural differences? Social situations? Physical tolerance for alcohol? Individual standards?

Conceptualization: The process of specifying what we mean by a term. In deductive research, conceptualization helps translate portions of an abstract theory into specific variables that can be used in testable hypotheses. In inductive research, conceptualization is an important part of the process used to make sense of related observations.

Many researchers now use the definition of *alcohol use disorder* (AUD) endorsed by the National Institute on Alcohol and Alcoholism (NIAAA), from the American Psychiatric Association's (2013) *Diagnostic and Statistical Manual of Mental Disorders (DSM-5):* "a chronic relapsing brain disease characterized by compulsive alcohol use, loss of control over alcohol intake, and a negative emotional state when not using." According to the NIAAA, an AUD is diagnosed according to the *DSM-5* standard, according to which at least 2 of 11 different criteria must have occurred during the same 12-month period. The criteria include "spent a lot of time drinking," "experienced craving," and "continued to drink even though it was causing trouble with your family or friends" (NIAAA n.d.). But, despite its popularity among professionals, we cannot judge the *DSM-5* definition of substance abuse as "correct" or "incorrect." Each researcher has the right to conceptualize as he or she sees fit. However, we can say that the *DSM-5* definition of substance abuse is useful, partly because it has been widely adopted. It is also stated in clear and precise language that minimizes differences in interpretation and maximizes understanding.

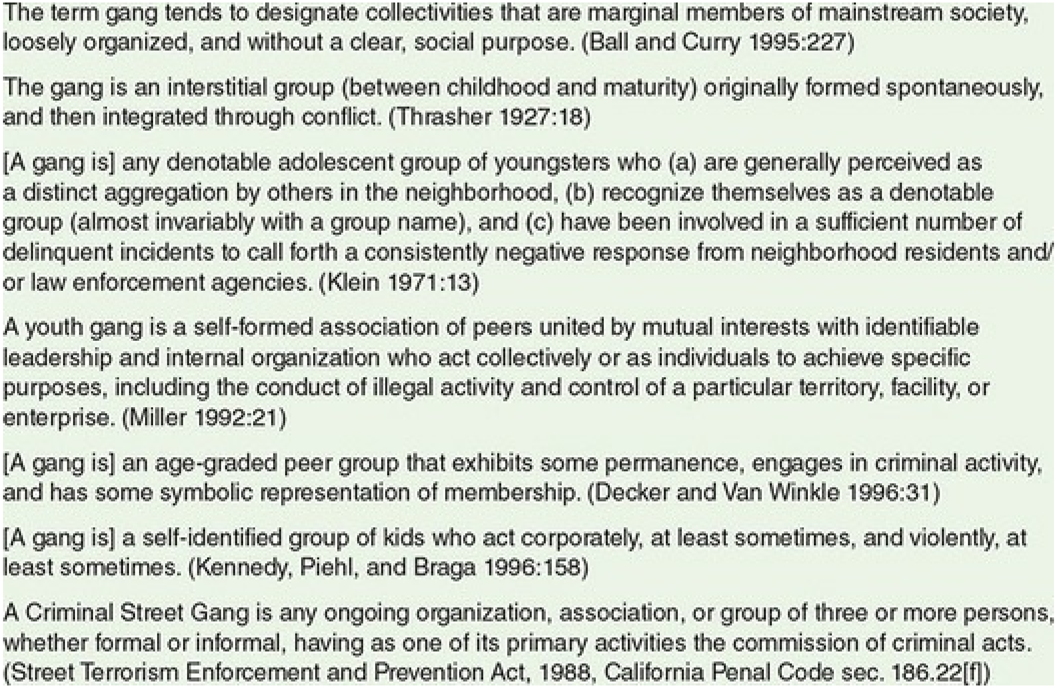
This clarity should not prevent us from recognizing that the definition reflects a particular theoretical orientation. *DSM-5* applies a medical *disease model* to mental illness (which is conceptualized, in *DSM-5,* to include substance use disorder). This theoretical model emphasizes behavioral and biological criteria instead of the social expectations that are emphasized in a social model of substance abuse. How we conceptualize reflects how we theorize.

Just as we can connect concepts to theory, we also can connect them to other concepts. What this means is that the definition of any one concept rests on a shared understanding of the other terms used in the definition. So if our audience does not already have a shared understanding of terms such as *significant adverse consequences* and *repeated use,* we must also define these terms before we are finished with the process of defining substance abuse.

Youth Gangs

Do you have a clear image in mind when you hear the term *youth gangs?* Although this is quite an ordinary term, social scientists' attempts to define precisely the concept, youth gang, have not yet succeeded: "Neither gang researchers nor law enforcement agencies can agree on a common definition . . . and a concerted national effort . . . failed to reach a consensus" (Howell 2003:75). Exhibit 4.1 lists a few of the many alternative definitions of youth gangs.

Exhibit 4.1 Alternative Definitions of Youth Gangs



The term gang tends to designate collectivities that are marginal members of mainstream society, loosely organized, and without a clear, social purpose. (Ball and Curry 1995:227)

The gang is an interstitial group (between childhood and maturity) originally formed spontaneously, and then integrated through conflict. (Thrasher 1927:18)

[A gang is] any denotable adolescent group of youngsters who (a) are generally perceived as a distinct aggregation by others in the neighborhood, (b) recognize themselves as a denotable group (almost invariably with a group name), and (c) have been involved in a sufficient number of delinquent incidents to call forth a consistently negative response from neighborhood residents and/or law enforcement agencies. (Klein 1971:13)

A youth gang is a self-formed association of peers united by mutual interests with identifiable leadership and internal organization who act collectively or as individuals to achieve specific purposes, including the conduct of illegal activity and control of a particular territory, facility, or enterprise. (Miller 1992:21)

[A gang is] an age-graded peer group that exhibits some permanence, engages in criminal activity, and has some symbolic representation of membership. (Decker and Van Winkle 1996:31)

[A gang is] a self-identified group of kids who act corporately, at least sometimes, and violently, at least sometimes. (Kennedy, Piehl, and Braga 1996:158)

A Criminal Street Gang is any ongoing organization, association, or group of three or more persons, whether formal or informal, having as one of its primary activities the commission of criminal acts. (Street Terrorism Enforcement and Prevention Act, 1988, California Penal Code sec. 186.22[f])

Source: Based on Howell, Preventing and Reducing Juvenile Delinquency, 2003:76 SAGE.

What is the basis of this conceptual difficulty? Researcher James Howell (2003:27–28) suggests that defining the term *youth gangs* has been difficult for four reasons:

1. Youth gangs are not particularly cohesive.
2. Individual gangs change their focus over time.
3. Many have a “hodgepodge of features,” with diverse members and unclear rules.
4. There are many incorrect but popular myths about youth gangs.

Which of the alternative definitions addresses these difficulties most effectively? Which makes the most sense to you?

Poverty

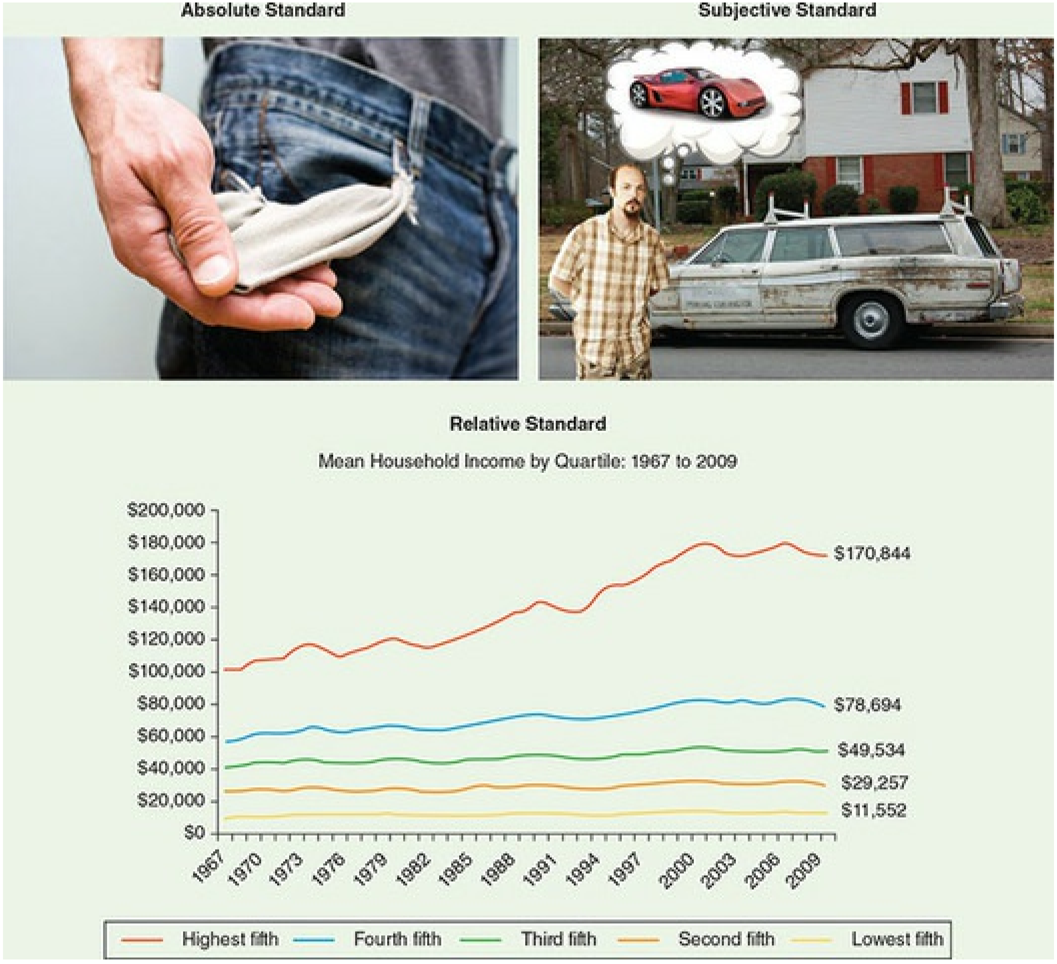
Decisions about how to define a concept reflect the theoretical framework that guides the researchers. For example, the concept *poverty* has always been somewhat controversial, because different conceptualizations of poverty lead to different estimates of its prevalence and different social policies for responding to it.

Most of the statistics about the U.S. poverty rate that you see in the news reflect a conception of poverty that was formalized by Mollie Orshansky of the Social Security Administration in 1965 and subsequently adopted by the federal government and many researchers (Putnam 1977). She defined *poverty* in terms of what is called an *absolute* standard, based on the amount of money required to purchase an emergency diet that is estimated to be nutritionally adequate for about 2 months. The idea is that people are truly poor if they can just barely purchase the food they need and other essential goods. This poverty standard is adjusted for household size and composition (number of children and adults), and the minimal amount of money needed for food is multiplied by three because a 1955 survey indicated that poor families spend about one third of their incomes on food (Orshansky 1977).

Does this sound straightforward? As is often the case with important concepts, the meaning of an absolute poverty standard has been the focus of a vigorous debate (Eckholm 2006:A8). Although the traditional definition of absolute poverty only accounts for a family’s cash income, some observers argue that noncash benefits that low-income people can receive, such as food stamps, housing subsidies, and tax rebates, should be added to cash income before the level of poverty is calculated. Douglas Besharov of the American Enterprise Institute terms this approach “a much needed corrective” (Eckholm 2006:A8). But some social scientists have proposed increasing the absolute standard for poverty so that it reflects what a low-income family must spend to maintain a “socially acceptable standard of living” that allows for a telephone, house repairs, and decent clothes (Uchitelle 1999). A new “Multidimensional Poverty Index” (MPI) to aid international comparisons considers absolute deprivations in health, education, and living standards (Alkire et al. 2011). Others argue that the persistence of poverty should be considered, so someone who is poor for no more than a year, for example, is distinguished from someone who is poor for many years (Walker, Tomlinson, and Williams 2010:367–368). Any change in the definition of poverty will change eligibility for government benefits such as food stamps and Medicaid, so the feelings about this concept run deep.

Some social scientists disagree altogether with the absolute standard and have instead urged adoption of a *relative* poverty standard (see Exhibit 4.2). They identify the poor as those in the lowest fifth or tenth of the income distribution or as those having some fraction of the average income. The idea behind this relative conception is that poverty should be defined in terms of what is normal in a given society at a particular time. “For example, while a car may be a luxury in some poor countries, in a country where most families own cars and public transportation is inadequate, a car is a basic necessity for finding and commuting to work” (Mayrl et al. 2004:10).

Exhibit 4.2 Absolute, Relative, and Subjective Poverty Standards



Source: Based on Giovanni Vecchi, Universita di Roma “Tor Vergata,” Poverty Lines. Bosnia and Herzegovina Poverty Analysis Workshop, September 17–21, 2007.

Some social scientists prefer yet another conception of poverty. With the *subjective* approach, *poverty* is defined as what people think would be the minimal income they need to make ends meet. Of course, many have argued that this approach is influenced too much by the different standards that people use to estimate what they “need” (Ruggles 1990:20–23). There is a parallel debate about the concept of “subjective well-being,” which is now measured annually in the United Kingdom by its Office of National Statistics with responses (on a 10-point scale) to four questions (Venkatapuram 2013:9):

1. Overall, how satisfied are you with your life nowadays?
2. Overall, to what extent do you feel the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?
4. Overall, how anxious did you feel yesterday?

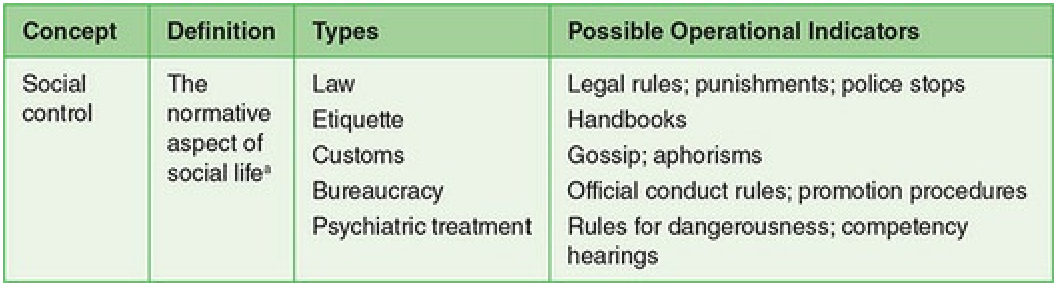
Which do you think is a more reasonable approach to defining poverty: some type of absolute standard, a relative standard, or a subjective standard? Be careful here: Conceptualization has consequences! Research using the standard absolute concept of poverty indicated that the percentage of Americans in poverty declined by 1.7% in the 1990s, but use of a relative concept of poverty led to the conclusion that poverty increased by 2.7% (Mayrl et al. 2004:10). No matter which conceptualization we decide to adopt, our understanding of the concept of poverty will be sharpened after we consider these alternative definitions.

From Concepts to Indicators

Identifying the concepts we will study, specifying dimensions of these concepts, and defining their meaning only begin the process of connecting our ideas to concrete observations. If we are to conduct empirical research involving a concept, we must be able to distinguish it in the world around us and determine how it may change over time or differ between persons or locations. Operationalization involves connecting concepts to measurement operations. You can think of it as the empirical counterpart of the process of conceptualization. When we conceptualize, we specify what we mean by a term (see Exhibit 4.3). When we operationalize, we identify specific measurements we will take to indicate that concept in empirical reality. Researchers also find that the process of figuring out how to measure a concept helps improve their understanding of what the concept means (Bartholomew 2010:457). Improving conceptualization and improving operationalization go hand in hand.

Operationalization: The process of specifying the measures that will indicate the value of cases on a variable.

Exhibit 4.3 Conceptualization and Operationalization of Social Control



Concept	Definition	Types	Possible Operational Indicators
Social control	The normative aspect of social life[a]	Law	Legal rules; punishments; police stops
		Etiquette	Handbooks
		Customs	Gossip; aphorisms
		Bureaucracy	Official conduct rules; promotion procedures
		Psychiatric treatment	Rules for dangerousness; competency hearings

Source: Based on Black, Donald. 1976. The Behavior of Law. New York: Academic Press.

Exhibit 4.3 illustrates conceptualization and operationalization by using the concept of *social control,* which Donald Black (1984) defined as "all of the processes by which people define and respond to deviant behavior" (p. xi). What observations can indicate this conceptualization of social control? Billboards that condemn drunk driving? Proportion of persons arrested in a community? Average length of sentences for crimes? Types of bystander reactions to public intoxication? Gossiping among neighbors? Some combination of these? Should we distinguish formal social control such as laws and police actions from informal types of social control such as social stigma? If we are to conduct research on the concept of social control, we must identify empirical indicators that are pertinent to our theoretical concerns.

Indicator: The question or other operation used to indicate the value of cases on a variable.

In the News Research in the News: Are Teenagers Replacing Drugs With Smartphones?

For Further Thought?

As high-school-aged teens' use of smartphones and tablets has accelerated in recent years, their use of illicit drugs other than marijuana has actually been dropping. Could the first trend be responsible to some extent for the second? Substance abuse expert Silvia Martins of Columbia University thinks this "is quite plausible." According to Nora Volkow, the director of the National Institute on Drug Abuse, "teens can get literally high when playing these [computer] games." Teens quoted in the article agreed, but other experts proposed other explanations. Professor James Anthony of Michigan State University admitted that "there is very little hard, definitive evidence on the subject."

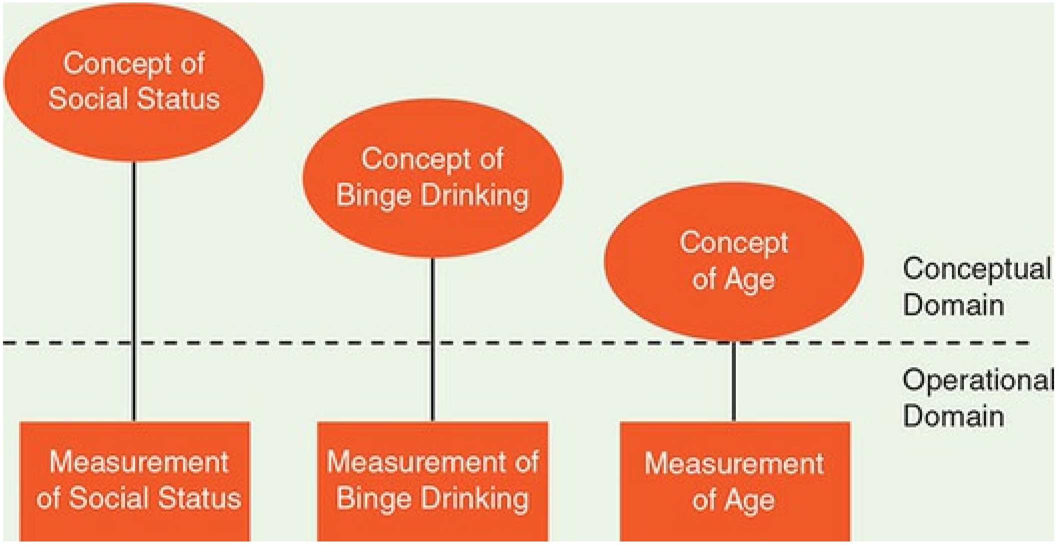
1. Should the concept of *addiction* be applied to behavior on modern technology devices? How would you define the concept of addiction?
2. Can we depend on self-report measures of drug (and technology) use? (The research described here used questions from the Monitoring the Future survey.) What measurement challenges can you think of?

News source: Richtel, Matt. 2017. "Are Teenagers Replacing Drugs With Smartphones?" *The New York Times,* March 13.

Abstract and Concrete Concepts

Concepts vary in their level of abstraction, and this, in turn, affects how readily we can specify the indicators pertaining to the concept. We may not think twice before we move from a conceptual definition of *age* as time elapsed since birth to the concrete indicator "years since birth." Binge drinking is also a relatively concrete concept, but it requires a bit more thought (see Exhibit 4.4). As you've seen, most researchers define *binge drinking* conceptually as heavy episodic drinking and operationally as drinking five or more drinks in a row (for men) (Wechsler et al. 2002:205). That's pretty straightforward, although we still need to specify the questions that will be used to determine the frequency of drinking. Jang et al. (2017), the subject of this chapter's "Research That Matters" feature, did not even bother giving a conceptual definition of "substance use behaviors." They just listed the specific questions they used to indicate these behaviors.

Exhibit 4.4 Varying Distances Between Concepts and Measures



Source: Adapted from Viswanathan 2005:7. Measurement Error and Research Design.

An abstract concept such as *social status* may have a clear role in social theory but still have different meanings in different social settings. Indicators that pertain to social status may include level of esteem in a group, extent of influence over others, level of income and education, or number of friends. It is very important to specify what we mean by an abstract concept such as social status in a particular study and to choose appropriate indicators to represent this meaning.

You have already learned in Chapter 2 that variables are phenomena that vary (and I hope you have practiced using the language of variables and hypotheses with the interactive exercises on the book's study site). Where do variables fit in the continuum from concepts to operational indicators that is represented in Exhibit 4.3? Think of it this way: Usually, the term *variable* is used to refer to some specific aspect of a concept that varies, and for which we then have to select even more concrete indicators. For example, research on the *concept* of social support might focus on the *variable,* "level of perceived support," and we might then select as our *indicator* the responses to a series of statements about social support, such as this one from S. Cohen et al.'s (1985) social support index, the "Interpersonal Support Evaluation List": "If I needed a quick emergency loan of $100, there is someone I could get it from" (p. 93). Identifying the variables we will measure is a necessary step on the road to developing our specific measurement procedures. I give more examples in the next section.

The term *variable* is sometimes used interchangeably with the term *indicator,* however, which means you might find "crime rate" or "importance of extrinsic rewards" being termed as either variables or indicators. Sometimes the term *variable* is used to refer to phenomena that are more abstract, such as "alienation" or "social capital." You might hear one researcher referring to *social support* as one of the important concepts in a study, another referring to it as a variable that was measured, and another calling it an indicator of group cohesion. The important thing to remember is that we need to define clearly the concepts we use and then develop specific procedures for identifying variation in the variables related to these concepts.

Bear in mind that concepts don't necessarily vary. For example, gender may be an important concept in a study of influences on binge drinking, but it isn't a variable in a study of members of a fraternity. When we explain excessive drinking in the fraternity, we might attach great importance to the all-male fraternity subculture. However, because gender doesn't vary in this setting, we won't be able to study differences in binge drinking between male and female students. So, gender will be a constant, not a variable, in this study (unless we expand our sample to include members of both sororities and fraternities, or perhaps the general student population).

How do we know what concepts to consider and then which variables to include in a study? It's very tempting, and all too common, to try simply to measure everything by including in a study every variable we can think of that might have something to do with our research question. This haphazard approach will inevitably result in the collection of some data that are useless and the failure to collect some data that are important. Instead, a careful researcher will examine relevant theories to identify key concepts, review prior research to learn how useful different indicators have been, and assess the resources available for measuring adequately the variables in the specific setting to be studied.

Operationalizing the Concept of Race

Race is an important concept in social research. In research applications as in everyday life, the concept of race is often treated as if it is an obvious distinction of several categories based on physical appearance (in turn related to ancestry). But in fact, what people mean by the concept of *race* has varied over time and differs between countries. These inconsistencies become clear only when we examine how the concept of race has been operationalized in specific questions.

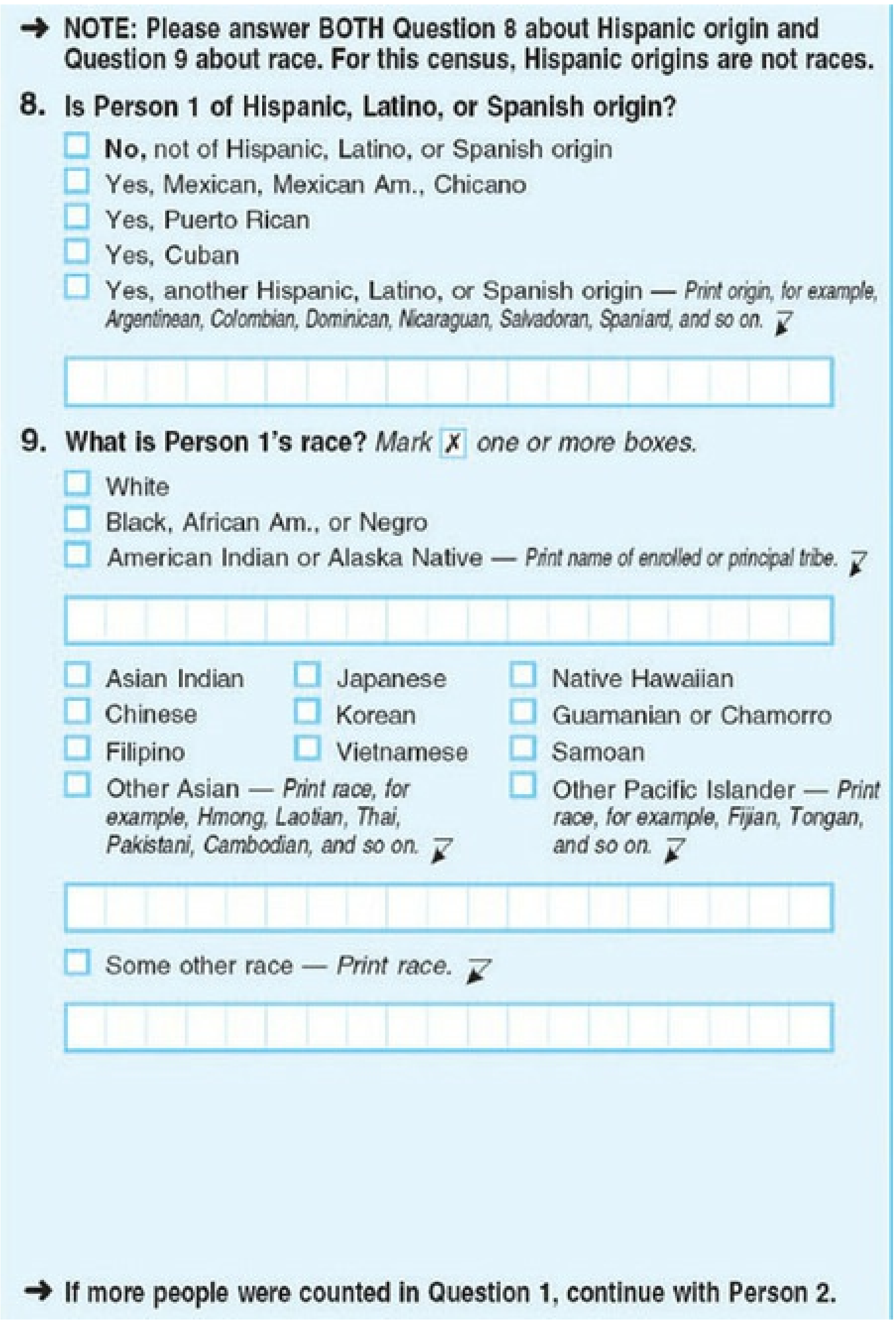
Repeated changes in questions about race in the decennial U.S. Census reflect social and political pressures (Snipp 2003:565–567). Race was not assessed directly in the censuses of 1790, 1800, or 1810, but slaves and American Indians were distinguished from others. In 1820, 1830, and 1840, *color* was distinguished; then, as racial consciousness heightened before the Civil War, the 1850 census added the category of *Mulatto* to identify people of mixed-race parentage. In response to concerns with increasing immigration, *Chinese* and *Indian* (Asian) persons were distinguished in 1860 and then *Japanese* was added to the list in 1870. In the 1890 census, *Octoroons* and *Quadroons* were distinguished as different categories of Mulatto. The infamous 1896 U.S. Supreme Court decision *Plessy v. Ferguson* reflected the victory of "Jim Crow" legislation in the South and defined as black any person who had as much as one black ancestor. (Homer Plessy was a Louisiana shoemaker whose skin was white, but he was told he could not ride in the "whites only" section of the train because he had one black ancestor.) By 1920, the U.S. Census reflected this absolute distinction between persons judged black and white by dropping the distinctions involving mixed-race ancestry. In 1930 and 1940, *Mexican* was distinguished, but in 1950, political pressure led to this category's being dropped as an ethnic minority; instead, Mexicans were treated as *white* (Snipp 2003:568–569).

By the late 1950s, the civil rights movement began to influence the concept of race as used by the U.S. Census. In 1960, the census shifted from assessing race on the basis of physical appearance to self-identification (Snipp 2003:569–570). As one result, the number of people identified as American Indians rose dramatically. More important, this shift reflected a change in thinking about the concept of race as reflecting primarily physical appearance and instead indicating cultural identification. In the 1970s, the Federal Interagency Committee on Education (FICE) established an ad hoc committee that led the 1980 U.S. Census (and all federal agencies) to use a five-category distinction: (1) American Indians and Alaska Natives, (2) Asians and Pacific Islanders, (3) non-Hispanic blacks, (4) non-Hispanic whites, and (5) Hispanics (Snipp 2003:572–574). In that census, Spanish/Hispanic origin or descent was asked as a question distinct from the question about race (U.S. Census Bureau 1981:3).

Constant: A number that has a fixed value in a given situation; a characteristic or value that does not change.

But the new concept of race reflected in these five categories only led to new complaints. Some parents in “mixed-race” marriages insisted that they should be able to classify their children as multiracial (Snipp 2003:575–576)—in opposition to some civil rights leaders concerned with diluting the numbers of Americans designated as black (Holmes 2001b:WK1). The chair of the Census Bureau’s Hispanic advisory committee complained, “We don’t fit into the categories that the Anglos want us to fit in” (Swarns 2004:A18). Sociologist Orlando Patterson (1997) and many others argued that the concept of ethnicity was more meaningful than race. As a result of these and other complaints, a new federal task force developed a new approach that allowed respondents to designate themselves as being of more than one race. The resulting question about race reflected these changes as well as increasing distinctions within what had been the Asians and Pacific Islanders category (see Exhibit 4.5). An official census report after the 2010 census also included this caveat about the definition:

Exhibit 4.5 The U.S. Census Bureau Ethnicity and Race Questions



→ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

☐ **No,** not of Hispanic, Latino, or Spanish origin
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark* ☒ *one or more boxes.*

☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

☐ Asian Indian
☐ Chinese
☐ Filipino
☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*

☐ Japanese
☐ Korean
☐ Vietnamese

☐ Native Hawaiian
☐ Guamanian or Chamorro
☐ Samoan
☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

☐ Some other race — *Print race.*

→ **If more people were counted in Question 1, continue with Person 2.**

Source: U.S. Census Bureau, 2010 Census Questionnaire.

> The race categories included in the census questionnaire generally reflect a social definition of race recognized in this country and are not an attempt to define race biologically, anthropologically, or genetically. In addition, it is recognized that the categories of the race question include race and national origin or sociocultural groups. (Humes, Jones, and Ramirez 2011:2).

With this new procedure in the 2010 U.S. Census, 36.7% of the country's 50.5 million

Latinos classified themselves as "some other race"—neither white nor black; they wrote in such terms as *Mayan, Tejano,* and *mestizo* to indicate their own preferred self-identification, using terms that focused on what social scientists term *ethnic* rather than on racial differences. In that same census, 3% of Americans identified themselves as multiracial (Humes et al. 2011). But this does not solve the problem of defining the concept of race. When David Harris and Jeremiah Sim (2002) analyzed responses in a national study of youths, they found that racial self-classification can vary with social context: 6.8% of youths classified themselves as multiracial when asked at school, but only 3.6% did so when asked at home. Even the presence of a parent during the in-home interview had an effect: Youths were less likely to self-identify as multiracial, rather than monoracial, in the presence of a parent. As already mentioned, more than a third of those who identified themselves as Hispanic declined to also classify themselves as white or black and instead checked "some other race." Some Arab American groups have now also asked for a special category (Vega 2014:A16).

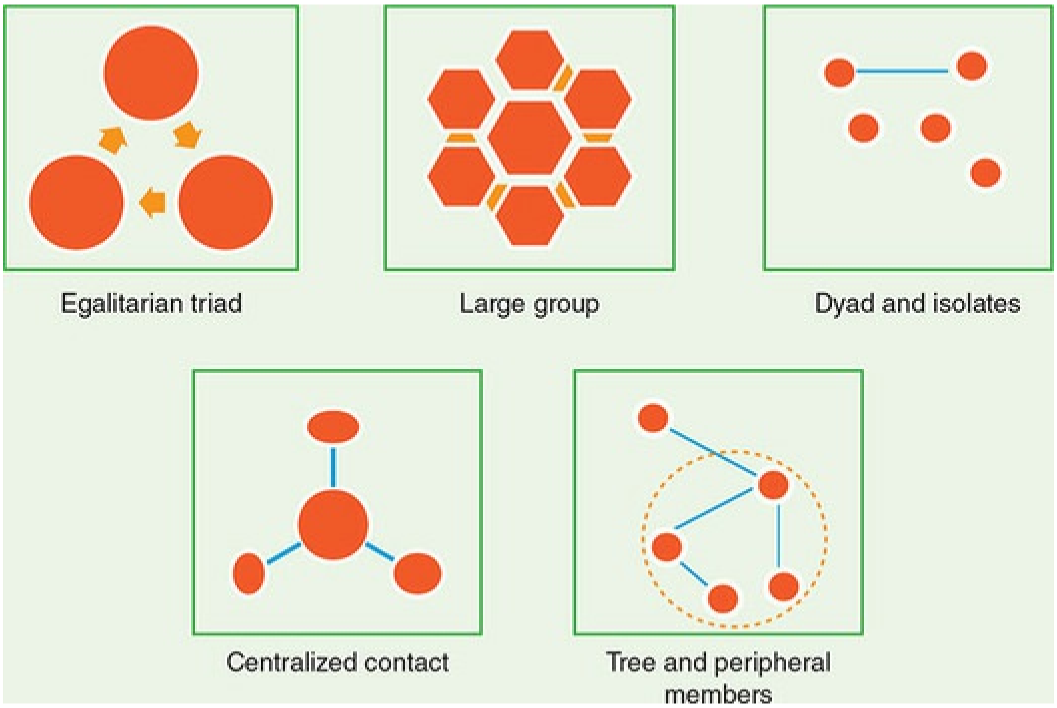
The concept of race also varies internationally, so any research involving persons in or from other countries may need to use a different definition of race. For example, Darryl Fears, former director of the Brazilian American Cultural Institute (in Washington, D.C.), explains how social conventions differ in Brazil: "In this country, if you are not quite white, then you are black." But in Brazil, "If you are not quite black, then you are white" (Fears 2002:A3). In Mexico, the primary ethnic distinction is between indigenous and nonindigenous residents, without a clear system of categorization based on skin color. Nonetheless, survey research indicates a marked preference for whiter skin and "profound social stratification by skin color" (Villarreal 2010:671). The conception and operationalization of race, then, varies with place.

Operationalizing Social Network Position

The concept of a social network has an intuitive appeal for sociological analysis because it focuses attention on the relationships between people that are the foundation of larger social structures (Scott 2013:5). Since Émile Durkheim's study of forms of solidarity, sociologists have been concerned with the nature of ties between people, how they differ between groups and societies, and how they affect social behavior. Social network analyses operationalize aspects of social network position through questions to respondents about who they are connected to in some (particular) way or through documents or other records that identify connections between people, organizations, or other entities. The social network analyst may then combine this information to create a picture of the overall structure of the network(s) under study.

For example, in their Teenage Health in Schools study of 3,146 adolescents in nine schools in western Scotland, Michael Pearson and his colleagues (2006) distinguished a number of social network positions, based on the adolescents' reports of their social connections. They distinguished some adolescents as participating in large or small groups, some who were isolated or participated in only a relationship dyad, and some as peripheral to others (Pearson et al. 2006:522). Exhibit 4.6 shows how some of these social network differences could be represented in diagrams, as well as how egalitarian and more hierarchical groups could be distinguished. A diagram like this can almost let you "feel the pain" of the adolescents who are peripheral or isolated compared with those who are involved in interlinked connections to their peers in social groups. In this way, social network analysis helps to understand individuals within their social context.

Exhibit 4.6 Social Network Positions



Source: Based on Pearson, Michael, Helen Sweeting, Patrick West, Robert Young, Jacki Gordon, and Katrina Turner. 2006. "Adolescent Substance Use in Different Social and Peer Contexts: A Social Network Analysis." Drugs: Education, Prevention and Policy 13:519–536.

One important finding was that youth in large groups were least likely to smoke or use drugs (Pearson et al. 2006:532). In another study of youth social networks, Susan Ennett and her colleagues (2006) found that youth who were more embedded in school social networks were less likely to abuse substances. Similarly, some social network position indicators are associated with better mental health (Wellman and Wortley 1990). Social network analysis has also been used to explain variation in behaviors ranging from delinquency, spreading of sexually transmitted infections, terrorism, and corporate crime to political donations and policy formulation (Yang et al. 2017).

From Observations to Concepts

Qualitative research projects usually take an inductive approach to the process of conceptualization. In an inductive approach, concepts emerge from the process of thinking about what has been observed, compared with the deductive approach that I have just described, in which we develop concepts on the basis of theory and then decide what should be observed to indicate that concept. So instead of deciding in advance which concepts are important for a study, what these concepts mean, and how they should be measured, if you take an inductive approach, you will begin by recording verbatim what you hear in intensive interviews or see during observational sessions. You will then review this material to identify important concepts and their meaning for participants. At this point, you may also identify relevant variables and develop procedures for indicating variation between participants and settings or variation over time. As your understanding of the participants and social processes develops, you may refine your concepts and modify your indicators. The sharp boundaries in quantitative research between developing measures, collecting data with those measures, and evaluating the measures often do not exist in inductive, qualitative research.

You will learn more about qualitative research in Chapter 10, but an example here will help you understand the qualitative measurement approach. For several months, Darin Weinberg (2000) observed participants in three drug abuse treatment programs in southern California. He was puzzled by the drug abuse treatment program participants' apparently contradictory beliefs—that drug abuse is a medical disease marked by "loss of control" but that participation in a therapeutic community can be an effective treatment. He discovered that treatment participants shared an "ecology of addiction" in which they conceived of being "in" the program as a protected environment, whereas being in the community was considered being "out there" in a place where drug use was inevitable—in "a space one's addiction compelled one to inhabit" (Weinberg 2000:609).

> I'm doin' real, real bad right now. . . . I'm havin' trouble right now staying clean for more than two days. . . . I hate myself for goin' out and I don't know if there's anything that can save me anymore. . . . I think I'm gonna die out there. (Weinberg 2000:609)

Participants contrasted their conscientiousness while in the program with the personal dissolution of those out in "the life."

So Weinberg developed the concepts of *in* and *out* inductively, in the course of the research, and he identified indicators of these concepts at the same time in the observational text. He continued to refine and evaluate the concepts throughout the research.

Conceptualization, operationalization, and validation were ongoing and interrelated processes. We'll study this process in more detail in Chapter 10.

Measurement

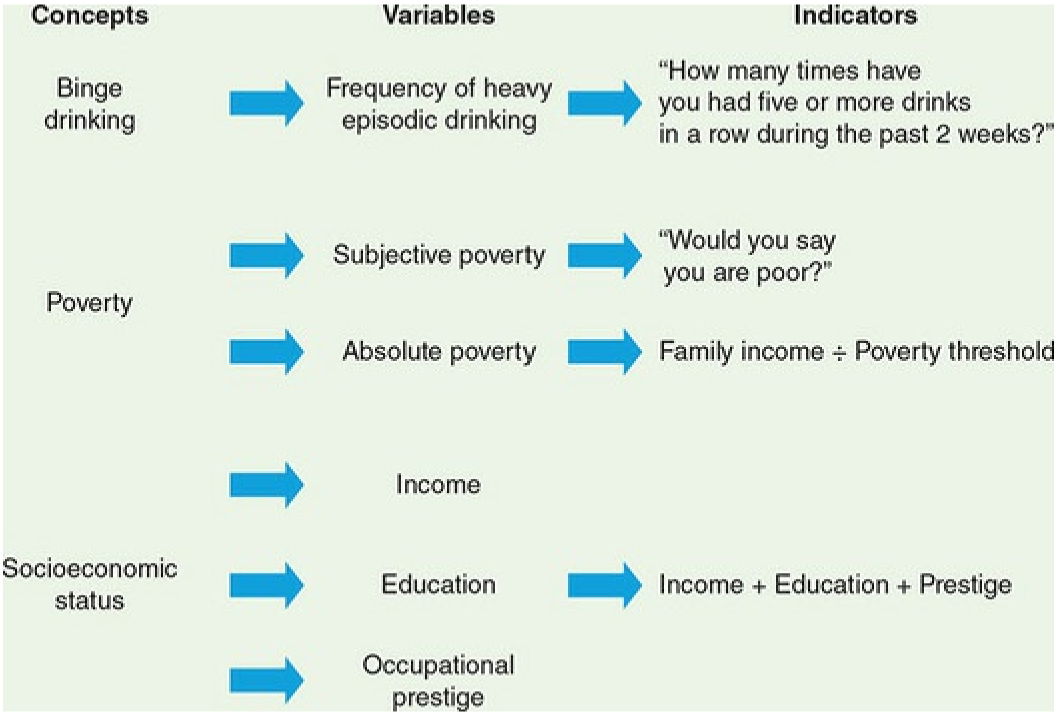
The deductive researcher proceeds from defining concepts in the abstract (conceptualizing) to identifying variables to measure, and finally to developing specific measurement procedures. Measurement is the "process of linking abstract concepts to empirical indicants" (Carmines and Zeller 1979:10). The goal is to achieve measurement validity, so the measures, or indicators, must actually measure the variables they are intended to measure.

Exhibit 4.7 represents the operationalization process in three studies. The first researcher defines her concept, *binge drinking,* and chooses one variable—frequency of heavy episodic drinking—to represent it. In the Monitoring the Future survey used in Jang et al.'s (2017) research, the single question, or indicator, used was "How many times have you had five or more drinks in a row during the past 2 weeks?" Binge drinking was defined as a response greater than 0. A researcher may define the concept *poverty* as having two aspects or dimensions: subjective poverty and absolute poverty. Subjective poverty is measured with responses to the survey question "Would you say you are poor?" Absolute poverty is measured by comparing family income to the poverty threshold. A researcher may operationalize the concept *socioeconomic status* by a position on a combination of responses to three measured variables: income, education, and occupational prestige.

Measurement: The process of linking abstract concepts to empirical indicants.

Social researchers have many options for operationalizing concepts. Measures can be based on activities as diverse as asking people questions, reading judicial opinions, observing social interactions, coding words in books, checking census data tapes, enumerating the contents of trash receptacles, or drawing urine and blood samples. Experimental researchers may operationalize a concept by manipulating its value. For example, to operationalize the concept of *exposure to antidrinking messages,* some subjects may listen to a talk about binge drinking while others do not. I will focus here on the operations of asking questions, observing behavior, using unobtrusive means of measuring people's behavior and attitudes, and using published data.

Exhibit 4.7 Concepts, Variables, and Indicators


Concepts
Variables
Indicators
Binge drinking
Frequency of heavy episodic drinking
"How many times have you had five or more drinks in a row during the past 2 weeks?"
Poverty
Subjective poverty
"Would you say you are poor?"
Absolute poverty
Family income ÷ Poverty threshold
Socioeconomic status
Income
Education
Occupational prestige
Income + Education + Prestige

The variables and particular measurement operations chosen for a study should be consistent with the research question. If we ask the evaluative research question "Are self-help groups more effective than hospital-based treatments in reducing drinking among substance abusers?" then we may operationalize "form of treatment" in terms of participation in these two types of treatments. However, if we are attempting to answer the explanatory research question "What influences the success of substance abuse treatment?" then we should probably consider what it is about these treatment alternatives that is associated with successful abstinence. Prior theory and research suggest that some of the important variables that differ between these treatment approaches are level of peer support, beliefs about the causes of alcoholism, and financial investment in the treatment.

Time and resource limitations must also be considered when we select variables and devise measurement operations. For many sociohistorical questions (e.g., "How has the poverty rate varied since 1950?"), census data or other published counts must be used. However, a historical question about the types of social bonds among combat troops in 20th-century wars probably requires retrospective interviews with surviving veterans. The validity of the data is lessened by the unavailability of many veterans from World War I and by problems of recall, but direct observation of their behavior during the war is certainly not an option.

Constructing Questions

Asking people questions is the most common and probably the most versatile operation for measuring social variables. Most concepts about individuals can be defined in such a way that measurement with one or more questions becomes an option. We associate questions with survey research, but questions are also often the basis of measures used in social experiments and in qualitative research. In this section, I introduce some options for writing single questions; in Chapter 8, I explain why single questions can be inadequate measures of some concepts, and then I examine measurement approaches that rely on multiple questions to measure a concept.

Of course, even though questions are, in principle, a straightforward and efficient means to measure individual characteristics, facts about events, level of knowledge, and opinions of any sort, they can easily result in misleading or inappropriate answers. Memories and perceptions of the events about which we might like to ask can be limited, and some respondents may intentionally give misleading answers. For these reasons, all questions proposed for a study must be screened carefully for their adherence to basic guidelines and then tested and revised until the researcher feels some confidence that they will be clear to the intended respondents and likely to measure the intended concept (Fowler 1995). Alternative measurement approaches will be needed when such confidence cannot be achieved.

Specific guidelines for reviewing survey questions are presented in Chapter 8; here, my focus is on the different types of questions used in social research.

Measuring variables with single questions is very popular. Public opinion polls based on answers to single questions are reported frequently in news articles and TV newscasts: “Do you favor or oppose U.S. policy . . . ?” “If you had to vote today, for which candidate would you vote?” Social science surveys also rely on single questions to measure many variables, for instance, “Overall, how satisfied are you with your job?” or “How would you rate your current health?”

Single questions can be designed with or without explicit response choices. The question that follows is a closed-ended (fixed-choice) question because respondents are offered explicit responses from which to choose. It has been selected from the Core Alcohol and Drug Survey distributed by the Core Institute, Southern Illinois University, for the Fund for the Improvement of Postsecondary Education (FIPSE) Core Analysis Grantee Group (Presley, Meilman, and Lyerla 1994):

> Compared to other campuses with which you are familiar, this campus’s use of alcohol is . . . (Mark one)

____ Greater than other campuses

____ Less than other campuses

____ About the same as other campuses

Most surveys of a large number of people contain primarily fixed-choice questions, which are easy to process with computers and analyze with statistics. With fixed-choice questions, respondents are also more likely to answer the questions that the researcher really wants them to answer. Including response choices reduces ambiguity and makes it easier for respondents to answer. However, fixed-response choices can obscure what people really think if the choices do not match the range of possible responses to the question; many studies show that some respondents will choose response choices that do not apply to them simply to give some sort of answer (Peterson 2000:39).

Closed-ended (fixed-choice) question: A survey question that provides preformatted response choices for the respondent to circle or check.

Most important, response choices should be mutually exclusive and exhaustive, so that every respondent can find one and only one choice that applies to him or her (unless the question is of the "Check all that apply" format). To make response choices exhaustive, researchers may need to offer at least one option with room for ambiguity. For example, a questionnaire asking college students to indicate their school status should not use freshman, sophomore, junior, senior, and graduate student as the only response choices. Most campuses also have students in a "special" category, so you might add "Other (please specify)" to the five fixed responses to this question. If respondents do not find a response option that corresponds to their answer to the question, they may skip the question entirely or choose a response option that does not indicate what they are really thinking.

Mutually exclusive: A question's response choices are mutually exclusive when every case can be classified as having only one attribute (or value).

Exhaustive: A question's response choices are exhaustive when they cover all possible responses.

Open-ended questions, questions without explicit response choices, to which respondents write in their answers, are preferable when the range of responses cannot adequately be anticipated—namely, questions that have not previously been used in surveys and questions that are asked of new groups. Open-ended questions can also lessen confusion about the meaning of responses involving complex concepts. The next question is an open-ended version of the earlier fixed-choice question:

> How would you say alcohol use on this campus compares with that on other campuses?

In qualitative research, open-ended questions are often used to explore the meaning respondents give to abstract concepts. *Mental illness,* for example, is a complex concept that tends to have different meanings for different people. In a survey I conducted in homeless shelters, I asked the staff members whether they believed that people at the shelter had become homeless because of mental illness (Schutt 1992). When given fixed-response choices, 47% chose "Agree" or "Strongly agree." However, when these same staff members were interviewed in depth, with open-ended questions, it became clear that the meaning of these responses varied among staff members. Some believed that mental illness caused homelessness by making people vulnerable in the face of bad luck and insufficient resources:

> Mental illness [is the cause]. Just watching them, my heart goes out to them. Whatever the circumstances were that were in their lives that led them to the streets and being homeless I see it as very sad. . . . Maybe the resources weren't there for them, or maybe they didn't have the capabilities to know when the resources were there. It is misfortune. (Schutt 1992:7)

Other staff believed that mental illness caused people to reject housing opportunities:

> I believe because of their mental illness that's why they are homeless. So for them to say I would rather live on the street than live in a house and have to pay rent, I mean that to me indicates that they are mentally ill. (Schutt 1992:7)

Just like fixed-choice questions, open-ended questions should be reviewed carefully for clarity before they are used. For example, if respondents are just asked, "When did you move to Boston?" they might respond with a wide range of answers: "In 1944." "After I had my first child." "When I was 10." "Twenty years ago." Such answers would be very hard to compare. To avoid ambiguity, rephrase the question to guide the answer in a certain direction, such as "In what year did you move to Boston?" or provide explicit response choices (Center for Survey Research 1987).

The decision to use closed-ended or open-ended questions can have important consequences for the information reported. Leaving an attitude or behavior off a fixed set of response choices is likely to mean that it is not reported, even if an "other" category is provided. However, any attitude or behavior is less likely to be reported if it must be volunteered in response to an open-ended question (Schwarz 2010:48).

Open-ended question: A survey question to which the respondent replies in his or her own words, either by writing or by talking.

Making Observations

Observations can be used to measure characteristics of individuals, events, and places. The observations may be the primary form of measurement in a study, or they may supplement measures obtained through questioning.

Direct observations can be used as indicators of some concepts. For example, Albert Reiss (1971a) studied police interaction with the public by riding in police squad cars, observing police–citizen interactions, and recording their characteristics on a form. Notations on the form indicated variables such as how many police–citizen contacts occurred, who initiated the contacts, how compliant citizens were with police directives, and whether police expressed hostility toward the citizens.

Using a different approach, psychologists Dore Butler and Florence Geis (1990) studied unconscious biases and stereotypes that they thought might hinder the advancement of women and minorities in work organizations. In one experiment, discussion groups of male and female students were observed from behind one-way mirrors as group leaders presented identical talks to each group. The trained observers (who were not told what the study was about) rated the number of frowns, furrowed brows, smiles, and nods of approval as the group leaders spoke. (The leaders themselves did not know what the study was about.) Group participants made disapproving expressions, such as frowns, more often when the group leader was a woman than when the leader was a man. To make matters worse, the more the women talked, the less attention they were given. Butler and Geis concluded that there was indeed a basis for discrimination in these unconscious biases.

Psychologists Joshua Correll, Bernadette Park, Charles Judd, and Bernd Wittenbrink (2002) used an even more creative approach to measure unconscious biases that could influence behavior despite an absence of conscious prejudice. Their approach focused on measuring reaction times to controlled observations. Correll et al. (2002) constructed a test in which individuals played a video game that required them to make a split-second decision of whether to shoot an image of a person who was holding what was a gun in some pictures and a nonlethal object such as a camera, cell phone, or bottle in others. In this ambiguous situation, white respondents were somewhat more likely to shoot a black man holding a nonlethal object than they were to shoot a white man holding a nonlethal object.

Observations may also supplement data collected in an interview study. This approach was used in a study of homeless persons participating in the Center for Mental Health Services' Access to Community Care and Effective Services and Supports (ACCESS) program. After a 47-question interview, interviewers were asked to record observations that would help indicate whether the respondent was suffering from a major mental illness. For example, the interviewers indicated, on a rating scale from 0 to 4, the degree to which the homeless participants appeared to be responding, during the interview, to voices or noises that others couldn’t hear or to other private experiences (U.S. Department of Health and Human Services 1995).

Direct observation is often the method of choice for measuring behavior in natural settings, as long as it is possible to make the requisite observations. Direct observation avoids the problems of poor recall and self-serving distortions that can occur with answers to survey questions. It also allows measurement in a context that is more natural than an interview. But observations can be distorted, too. Observers do not see or hear everything, and what they do see is filtered by their own senses and perspectives. Disagreements about crowd size among protestors, police, and journalists are notorious, even though there is a good method of estimating crowd size based on the “carrying capacity” of public spaces (McPhail and McCarthy 2004). When the goal is to observe behavior, measurement can be distorted because the presence of an observer may cause people to act differently than they would otherwise (Emerson 1983). I discuss these issues in more depth in Chapters 10 and 11, but it is important to consider them whenever you read about observational measures.

Collecting Unobtrusive Measures

Unobtrusive measures allow us to collect data about individuals or groups without their direct knowledge or participation. In their classic book (now revised), Eugene Webb and his colleagues (2000) identified four types of unobtrusive measures: physical trace evidence, archives (available data), simple observation, and contrived observation (using hidden recording hardware or manipulation to elicit a response). We have already considered observational data and we will consider available data (from "archives") in the next section, so I focus here on the other approaches suggested by Webb et al. (2000).

Unobtrusive measure: A measurement based on physical traces or other data that are collected without the knowledge or participation of the individuals or groups that generated the data.

The physical traces of past behavior are one type of unobtrusive measure that is most useful when the behavior of interest cannot be observed directly (perhaps because it is hidden or occurred in the past) and has not been recorded in a source of available data. To measure the prevalence of drinking in college dorms or fraternity houses, we might count the number of empty bottles of alcoholic beverages in the surrounding dumpsters. Student interest in the college courses they are taking might be measured by counting the number of times that books left on reserve as optional reading are checked out or by the number of class handouts left in trash barrels outside a lecture hall. Webb and his colleagues (2000:37) suggested measuring the interest in museum exhibits by the frequency with which tiles in front of the exhibits needed to be replaced. Social variables can also be measured by observing clothing, hair length, or people's reactions to such stimuli as dropped letters or jaywalkers.

You can probably see that care must be taken to develop trace measures that are useful for comparative purposes. For instance, comparison of the number of empty bottles in dumpsters outside different dorms can be misleading; at the very least, you would need to account for the number of residents in the dorms, the time since the last trash collection, and the accessibility of each dumpster to passersby. Counts of usage of books on reserve will be useful only if you consider how many copies of the books are on reserve for the course, how many students are enrolled in the course, and whether reserve reading is required. Measures of tile erosion in the museum must account for the nearness of each exhibit to doors, other popular exhibits, and so on (Webb et al. 2000:47–48).

Using Available Data

Government reports are rich and readily accessible sources of social science data. Organizations ranging from nonprofit service groups to private businesses also compile a wealth of figures that may be available to social scientists for some purposes. In addition, the data collected in many social science surveys are archived and made available for researchers who were not involved in the original survey project.

Before we assume that available data will be useful, we must consider how appropriate they are for our concepts of interest. We may conclude that some other measure would provide a better fit with a concept or that a particular concept simply cannot be adequately operationalized with the available data. For example, law enforcement and health statistics provide several community-level indicators of substance abuse (Gruenewald et al. 1997). Statistics on arrests for the sale and possession of drugs, drunk driving arrests, and liquor law violations (such as sales to minors) can usually be obtained on an annual basis, and often quarterly, from local police departments or state crime information centers. Health-related indicators of substance abuse at the community level include single-vehicle fatal crashes, the rate of mortality from alcohol or drug abuse, and the use of alcohol and drug treatment services.

Indicators such as these cannot be compared across communities or over time without reviewing carefully how they were constructed. The level of alcohol in the blood that is legally required to establish intoxication can vary among communities, creating the appearance of different rates of substance abuse even though drinking and driving practices may be identical. Enforcement practices can vary among police jurisdictions and over time (Gruenewald et al. 1997:14).

We also cannot assume that available data are accurate, even when they appear to measure the concept in which we are interested in a way that is consistent across communities. “Official” counts of homeless persons have been notoriously unreliable because of the difficulty in locating homeless persons on the streets, and government agencies have, at times, resorted to “guesstimates” by service providers (Rossi 1989). Even available data for such seemingly straightforward measures as counts of organizations can contain a surprising amount of error. For example, a 1990 national church directory reported 128 churches in a midwestern U.S. county; an intensive search in that county in 1992 located 172 churches (Hadaway, Marler, and Chaves 1993:744). Perhaps 30% or 40% of death certificates identify incorrectly the cause of death (Altman 1998).

Government statistics that are generated through a central agency such as the U.S. Census Bureau are often of high quality, but caution is warranted when using official data collected by local levels of government. For example, the Uniform Crime Reports (UCR) program administered by the Federal Bureau of Investigation (FBI) imposes standard classification criteria, with explicit guidelines and regular training at the local level, but data are still inconsistent for many crimes. Consider only a few of the many sources of inconsistency between jurisdictions: Variation in the classification of forcible rape cases due to differences in what is considered to be "carnal knowledge of a female"; different decisions about what is considered "more than necessary force" in the definition of "strong-arm" robberies; whether offenses in which threats were made but no physical injury occurred are classified as aggravated or simple assaults (Mosher, Miethe, and Phillips 2002:66). The National Incident-Based Reporting System (NIBRS) was designed to correct some of the problems with the UCR, but it is itself now being revised (Federal Bureau of Investigation n.d.; Mosher et al. 2002:70).

In some cases, problems with an available indicator can be lessened by selecting a more precise indicator. For example, the number of single-vehicle nighttime crashes, whether fatal or not, is a more specific indicator of the frequency of drinking and driving than is just the number of single-vehicle fatal accidents (Gruenewald et al. 1997:40–41). Focusing on a different level of aggregation may also improve data quality, because procedures for data collection may differ between cities, counties, states, and so on (Gruenewald et al. 1997:40–41). Only after factors such as legal standards, enforcement practices, and measurement procedures have been accounted for do comparisons between communities become credible.

Coding Content

Unobtrusive measures can also be created from diverse forms of media such as newspaper archives or magazine articles, TV or radio talk shows, legal opinions, historical documents, personal letters, or e-mail messages. Qualitative researchers may read and evaluate text, as Sally Lindsay and her colleagues (2007) did in their study of computer-mediated social support for people with diabetes (see Chapter 1). Quantitative researchers use content analysis to measure aspects of media such as the frequency of use of particular words or ideas or the consistency with which authors convey a particular message in their stories. An investigation of the drinking climate on campuses might include a count of the amount of space devoted to ads for alcoholic beverages in a sample of issues of the student newspaper. Campus publications also might be coded to indicate the number of times that statements discouraging substance abuse appear. With this tool, you could measure the frequency of articles reporting substance abuse–related crimes, the degree of approval of drinking expressed in TV shows or songs, or the relationship between region of the country and the amount of space devoted in the print media to drug usage.

Taking Pictures

Photographs record individual characteristics and social events, so they can become an important tool for investigating the social world. In recent years, photography has become a much more common part of the social world, as cameras embedded in cell phones and the use of websites and social media encourage taking and sharing photos. Sociologists and other social scientists are increasingly using photos as indicators of peoples' orientations in other times and places, as clues to the perspectives of the photographers themselves (Tinkler 2013:15).

Exhibit 4.8 displays a photo of two Ukrainian women embroidering traditional Ukrainian Easter towels, but they are doing so while sitting on a park bench in Italy where they are domestic live-in workers. Olena Fedyuk (2012) included the picture in her analysis of photos exchanged by Ukrainian domestic workers in Italy with their family members who remained in Ukraine. Fedyuk concludes that pictures like these help convey the reassuring message to their families that these women are still focused on their home country and family and are not engaged in Italian social life.

Combining Measurement Operations

Asking questions, making observations, using unobtrusive indicators, including available data or coding content, and taking pictures are interrelated measurement tools, each of which may include or be supplemented by the others. From people's answers to survey questions, the U.S. Census Bureau develops widely consulted census reports containing available data on people, firms, and geographic units in the United States. Data from employee surveys may be supplemented by information available in company records. Interviewers may record observations about those whom they question. Researchers may use insights gleaned from questioning participants to make sense of the social interaction they have observed. Unobtrusive indicators can be used to evaluate the honesty of survey respondents.

The available resources and opportunities often determine the choice of a particular measurement method, but measurement is improved if this choice also accounts for the particular concept or concepts to be measured. Responses to questions such as "How socially engaged were you at the party?" or "How many days did you use sick leave last year?" are unlikely to provide information as valid as, respectively, direct observation or company records. However, observations at social gatherings may not answer our questions about why some people do not participate; we may have to ask people. Or, if no record is kept of sick leaves in a company, we may have to ask direct questions.

Exhibit 4.8 Ukrainian Domestic Workers

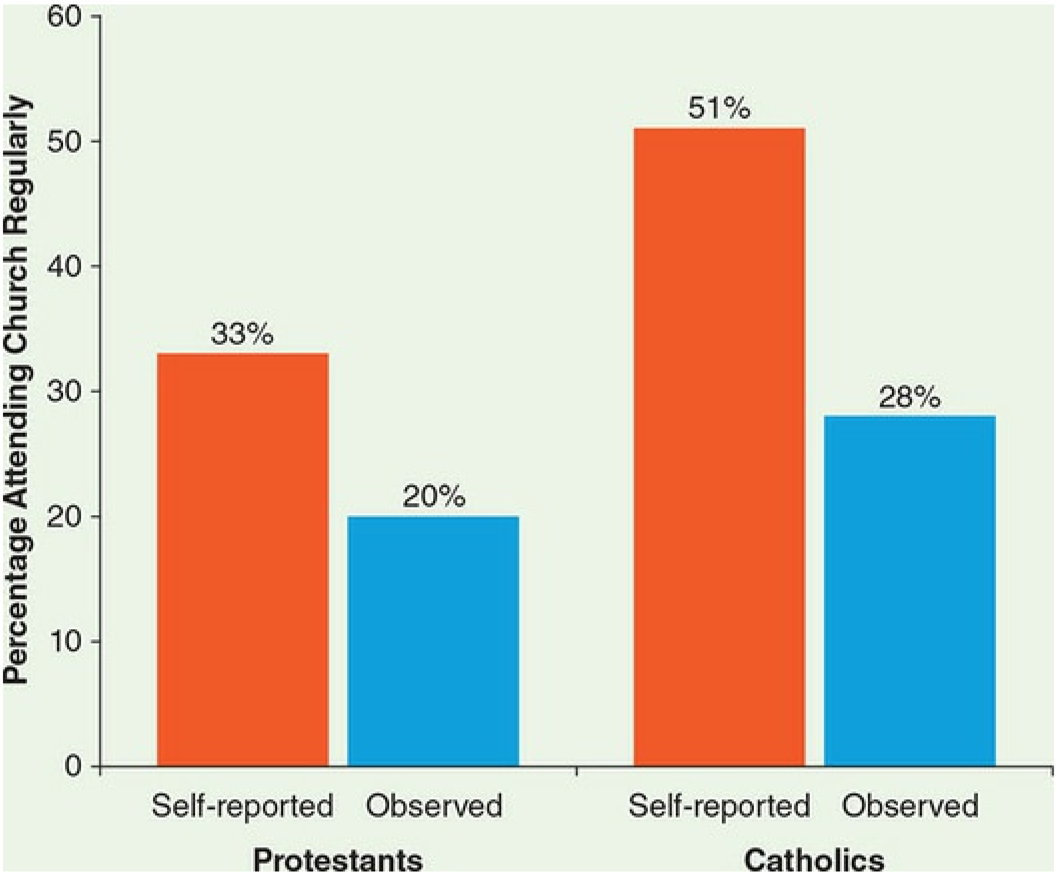
Source: Oksana Pronyuk. From Fedyuk, Olena. 2012. “Images of Transnational Motherhood: The Role of Photographs in Measuring Time and Maintaining Connections Between Ukraine and Italy.” Journal of Ethnic and Migration Studies 38:279–300.

Questioning can be a particularly poor approach for measuring behaviors that are socially desirable, such as voting or attending church, or that are socially stigmatized or illegal, such as abusing alcohol or drugs. The tendency of people to answer questions in socially approved ways was demonstrated in a study of church attendance in the United States (Hadaway et al. 1993). More than 40% of adult Americans say in surveys that they attend church weekly—a percentage much higher than in Canada, Australia, or Europe. However, a comparison of observed church attendance with self-reported attendance suggested that the actual rate of church attendance was much lower (see Exhibit 4.9). Always consider the possibility of measurement error when only one type of operation has been used. Of course, it is much easier to recognize this possibility than it is to determine the extent of error resulting from a particular measurement procedure. Refer to the February 1998 issue of the *American Sociological Review* for a fascinating exchange of views and evidence on the subject of measuring church attendance.

Triangulation—the use of two or more different measures of the same variable—can strengthen measurement considerably (Brewer and Hunter 1989:17). When we achieve similar results with different measures of the same variable, particularly when they are based on such different methods as survey questions and field-based observations, we can be more confident in the validity of each measure. If results diverge with different measures, it may indicate that one or more of these measures are influenced by more measurement error than we can tolerate. Divergence between measures could also indicate that they actually operationalize different concepts. An interesting example of this interpretation of divergent results comes from research on crime. Official crime statistics indicate only those crimes that are reported to and recorded by the police; when surveys are used to measure crimes with self-reports of victims, many “personal annoyances” are included as if they were crimes (Levine 1976).

Triangulation: The use of multiple methods to study one research question; also used to mean the use of two or more different measures of the same variable.

Exhibit 4.9 The Inadequacy of Self-Reports Regarding Socially Desirable Behavior: Observed Versus Self-Reported Church Attendance



Source: Data from Kirk C. Hadaway, Penny Long Marker, and Mark Chaves. 1993. “What the Polls Don’t Show: A Closer Look at U.S. Church Attendance.” American Sociological Review 58(6):741–752.

Careers and Research

Camila Mejia, Market Researcher

Camila Mejia majored in psychology and earned a graduate degree in clinical psychology at Universidad Pontificia Bolivariana in Colombia. After graduating, she started working in a somewhat unexplored field for psychology in Colombia: market research from a consumer psychology perspective. However, her experience reinforced her belief that we can't understand human behavior without taking account of the social world, and so with this thought in mind and her passion for research, she applied to the master's program in applied sociology at the University of Massachusetts Boston.

After earning her MA, Mejia returned to Colombia and began a new position in market research for a consumer goods company. Now she conducts social research to provide brands with information that can be used to understand consumer behavior. Her projects use social research methods ranging from ethnography and focus groups to in-depth interviews and surveys. She has been particularly impressed with the ability of ethnographic methods to engage with people in their daily lives in order to understand their buying habits. She is invited into homes to observe usage of consumer goods and accompanies consumers to different stores to observe what they buy, how they interact with brands, and their relationship with the marketplace. Her advice to new students is to seek innovative ways to apply sociology and research methods to understanding the social world.

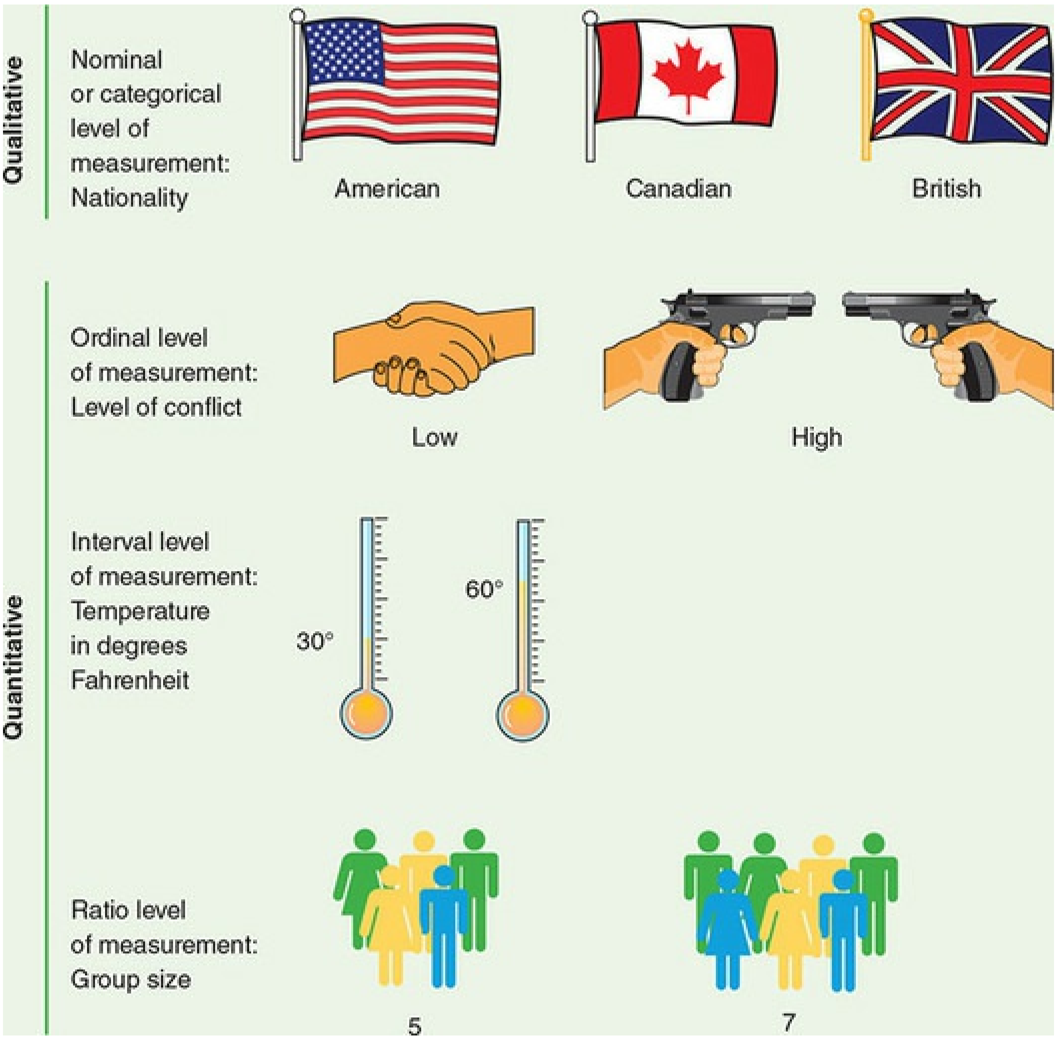
Levels of Measurement

Can you name the variables represented in Exhibit 4.9? One variable is “religion”; it is represented by only two attributes, or categories, in Exhibit 4.9—Protestant and Catholic—but you know that there are many others. You also know that one religion is not “more religion” than another; they are different in kind, not amount. The other variable represented in Exhibit 4.9 is “frequency of church attendance.” Of course, frequencies *do* differ in amount. We can say that religion—a qualitative indicator—and frequency of church attendance—a quantitative indicator—differ in their levels of measurement.

When we know a variable’s level of measurement, we understand more about how cases vary on that variable and so appreciate more fully what we have measured. Level of measurement also has important implications for the type of statistics that can be used with the variable, as you will learn in Chapter 14. There are four levels of measurement: (1) nominal, (2) ordinal, (3) interval, and (4) ratio. For most purposes, variables measured at the interval and ratio levels are treated in the same way, so I will sometimes refer to these two levels together as *interval–ratio.* Exhibit 4.10 depicts the differences among these four levels.

> Level of measurement: The mathematical precision with which the values of a variable can be expressed. The nominal level of measurement, which is qualitative, has no mathematical interpretation; the quantitative levels of measurement—ordinal, interval, and ratio—are progressively more precise mathematically.

Exhibit 4.10 Levels of Measurement


Qualitative
Nominal or categorical level of measurement: Nationality
American
Canadian
British
Quantitative
Ordinal level of measurement: Level of conflict
Low
High
Interval level of measurement: Temperature in degrees Fahrenheit
30°
60°
Ratio level of measurement: Group size
5
7

Nominal Level of Measurement

The nominal level of measurement identifies variables whose values have no mathematical interpretation; they vary in kind or quality but not in amount (they may also be called *categorical* or *qualitative variables*). It is conventional to refer to the values of nominal variables as *attributes* instead of values. The Jang et al. (2017:5) study provides an example in the question asked about marital status. Response options were provided as 1 = married, 2 = engaged, 3 = separated/divorced, 4 = widowed, and 5 = single. The numbers represent the particular response options, but these numbers do not tell us anything about the difference between the types of marital status except that they are different. "Engaged" is not one unit less of *marital status* than "separated/divorced." Nationality, occupation, religious affiliation, and region of the country are also measured at the nominal level. A person may be Spanish or Portuguese, but one nationality does not represent more nationality than another—just a different nationality (see Exhibit 4.10). A person may be a doctor or a truck driver, but one does not represent more occupation than the other. Of course, people may identify more strongly with one nationality than another, or one occupation may have a higher average income than another, but these are comparisons involving the variables, "strength of national identification" and "average income," not *nationality* or *occupation per se.*

Nominal level of measurement: Variables whose values have no mathematical interpretation; they vary in kind or quality, but not in amount.

Although the attributes of categorical variables do not have a mathematical meaning, they must be assigned to cases with great care. The attributes we use to measure, or to categorize, cases must be mutually exclusive and exhaustive:

- A variable's attributes or values are mutually exclusive if every case can have only one attribute.
- A variable's attributes or values are exhaustive when every case can be classified into one of the categories.

When a variable's attributes are mutually exclusive and exhaustive, every case corresponds to one, and only one, attribute.

I know this sounds pretty straightforward, and in many cases it is. However, what we think of as mutually exclusive and exhaustive categories may really be so only because of social convention; when these conventions change, or if they differ between the societies in a multicountry study, appropriate classification at the nominal level can become much more complicated. You learned of complexities such as this in the earlier discussion of the history of measuring race.

Issues similar to these highlight the importance of informed selection of concepts, careful conceptualization of what we mean by a term, and systematic operationalization of the procedures for indicating the attributes of actual cases. The debate regarding the concept of race also reminds us of the value of qualitative research that seeks to learn about the meaning that people give to terms, without requiring that respondents use predetermined categories.

Ordinal Level of Measurement

The first of the three quantitative levels is the ordinal level of measurement. At this level, the numbers assigned to cases specify only the order of the cases, permitting *greater than* and *less than* distinctions. The properties of variables measured at the ordinal level are illustrated in Exhibit 4.10 by the contrast between the levels of conflict in two groups. The first group, symbolized by two people shaking hands, has a low level of conflict. The second group, symbolized by two persons pointing guns at each other, has a high level of conflict. To measure conflict, we would put the groups "in order" by assigning the number 1 to the low-conflict group and the number 2 to the high-conflict group. The numbers thus indicate only the relative position or order of the cases. Although low level of conflict is represented by the number 1, it is not one less unit of conflict than high level of conflict, which is represented by the number 2.

The Favorable Attitudes Toward Antisocial Behavior Scale measures attitudes toward antisocial behavior among high school students with a series of questions that each involves an ordinal distinction (see Exhibit 4.11). The response choices to each question range from "very wrong" to "not wrong at all." There's no particular quantity of "wrongness" that these distinctions reflect, but the idea is that a student who responds that it is "not wrong at all" to a question about taking a handgun to school has a more favorable attitude toward antisocial behavior than does a student who says it is "a little bit wrong," which is in turn more favorable than those who respond "wrong" or "very wrong."

Ordinal level of measurement: A measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.

As with nominal variables, the different values of a variable measured at the ordinal level must be mutually exclusive and exhaustive. They must cover the range of observed values and allow each case to be assigned no more than one value. Often, questions that use an ordinal level of measurement simply ask respondents to rate their response to some question or statement along a continuum of, for example, strength of agreement, level of importance, or relative frequency. Like variables measured at the nominal level, variables measured at the ordinal level in this way classify cases in discrete categories and so are termed discrete measures.

A series of similar questions may be used instead of one question to measure the same concept. The set of questions in the Favorable Attitudes Toward Antisocial Behavior Scale in Exhibit 4.11 is a good example. In such a multi-item index, or *scale,* numbers are assigned to reflect the order of the responses (such as 1 for "very wrong," 2 for "wrong," 3 for “a little bit wrong,” and 4 for “not wrong at all”); these responses are then summed or averaged to create the index score. One person’s responses to the five questions in Exhibit 4.11 could thus range from 5 (meaning they said each behavior is “very wrong”) to 20 (meaning they said each behavior is “not wrong at all”). However, even though these are numeric scores, they still reflect an ordinal level of measurement because the responses they are based on involve only ordinal distinctions.

Interval Level of Measurement

The numbers indicating the values of a variable at the interval level of measurement represent fixed measurement units but have no absolute, or fixed, zero point. This level of measurement is represented in Exhibit 4.10 by the difference between two Fahrenheit temperatures. Although 60 degrees is 30 degrees hotter than 30 degrees, 60, in this case, is not twice as hot as 30. Why not? Because "heat" does not begin at 0 degrees on the Fahrenheit scale.

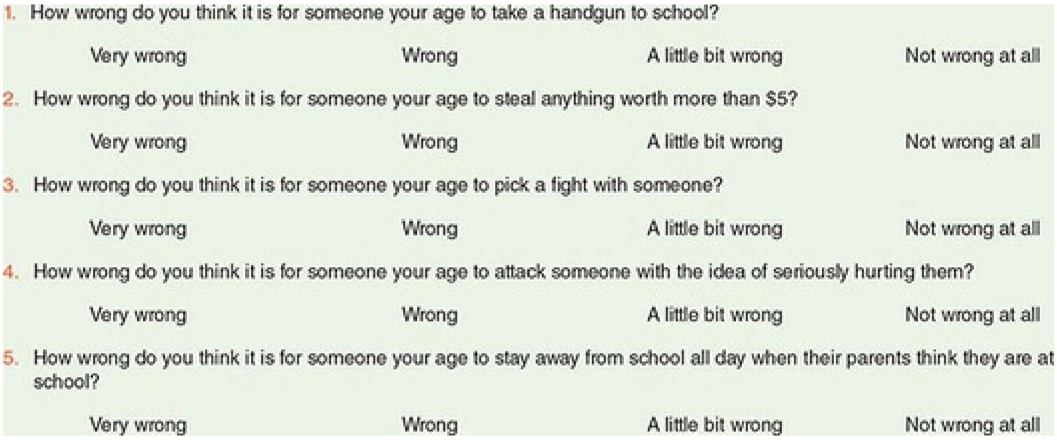
Discrete measure: A measure that classifies cases in distinct categories.

Index: The sum or average of responses to a set of questions about a concept.

Interval level of measurement: A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.

An interval-level measure is created by a scale that has fixed measurement units but no absolute, or fixed, zero point. The numbers can, therefore, be added and subtracted, but ratios are not meaningful. Again, the values must be mutually exclusive and exhaustive. There are few true interval-level measures in the social sciences, but cross-disciplinary investigations such as those that examine the linkage between climate and social life or historical change in social organization may involve interval measures such as temperature or calendar year.

Exhibit 4.11 Example of Ordinal Measures: Favorable Attitudes Toward Antisocial Behavior Scale



1. How wrong do you think it is for someone your age to take a handgun to school?

   Very wrong Wrong A little bit wrong Not wrong at all

2. How wrong do you think it is for someone your age to steal anything worth more than $5?

   Very wrong Wrong A little bit wrong Not wrong at all

3. How wrong do you think it is for someone your age to pick a fight with someone?

   Very wrong Wrong A little bit wrong Not wrong at all

4. How wrong do you think it is for someone your age to attack someone with the idea of seriously hurting them?

   Very wrong Wrong A little bit wrong Not wrong at all

5. How wrong do you think it is for someone your age to stay away from school all day when their parents think they are at school?

   Very wrong Wrong A little bit wrong Not wrong at all

Sources: Lewis, Chandra, Gwen Hyatt, Keith Lafortune, and Jennifer Lembach. 2010. History of the Use of Risk and Protective Factors in Washington State's Healthy Youth Survey. Portland, OR: RMC Research Corporation. Retrieved May 11, 2014, from https://www.askhys.net/library/Old/RPHistory.pdf, page 26. See also Arthur, Michael W., John S. Briney, J. David Hawkins, Robert D. Abbott, Blair L. Brooke-Weiss, and Richard F. Catalano. 2007. “Measuring Risk and Protection in Communities Using the Communities That Care Youth Survey.” Evaluation and Program Planning 30:197–211.

Many social scientists use indexes created by combining responses to a series of variables measured at the ordinal level as if they were interval-level measures. You have already seen an example of such a multi-item index in Exhibit 4.11. Scores on scales that are standardized in a distribution can also be treated as measured at the interval level. For example, the so-called Intelligence Quotient, or IQ score, is standardized so that a score of 100 indicates that a person is in the middle of the distribution, and each score above and below 100 indicates fixed points on the distribution of IQ scores in the population. A score of 114 would represent a score higher than that of 84% of the population.

Ratio Level of Measurement

The numbers indicating the values of a variable at the ratio level of measurement represent fixed measuring units and an absolute zero point (zero means absolutely no amount of whatever the variable indicates). For example, the following question was used on the National Minority SA/HIV Prevention Initiative Youth Questionnaire to measure number of days during the past 30 days that the respondent drank at least one alcoholic beverage. We can easily calculate the number of days that separate any response from any other response (except for the missing value of "don't know").

> During the past 30 days, on how many days did you drink one or more drinks of an alcoholic beverage?
>
> ☐ 0 days ☐ 12 days ☐ 24 days
>
> ☐ 1 day ☐ 13 days ☐ 25 days
>
> ☐ 2 days ☐ 14 days ☐ 26 days
>
> ☐ 3 days ☐ 15 days ☐ 27 days
>
> ☐ 4 days ☐ 16 days ☐ 28 days
>
> ☐ 5 days ☐ 17 days ☐ 29 days
>
> ☐ 6 days ☐ 18 days ☐ 30 days
>
> ☐ 7 days ☐ 19 days ☐ Don't know
>
> ☐ 8 days ☐ 20 days or can't say
>
> ☐ 9 days ☐ 21 days
>
> ☐ 10 days ☐ 22 days
>
> ☐ 11 days ☐ 23 days

Exhibit 4.10 also displays an example of a variable measured at the ratio level. The number of people in the first group is five, and the number in the second group is seven. The ratio of the two groups' sizes is then 1.4, a number that mirrors the relationship between the sizes of the groups. Note that there does not actually have to be any group with a size of zero; what is important is that the numbering scheme begins at an absolute zero—in this case, the absence of any people.

Ratio level of measurement: A measurement of a variable in which the numbers indicating a variable's values represent fixed measuring units and an absolute zero point.

For most statistical analyses in social science research, the interval and ratio levels of measurement can be treated as equivalent. In addition to having numerical values, both the interval and ratio levels also involve continuous measures: The numbers indicating the values of variables are points on a continuum, not discrete categories. But despite these similarities, there is an important difference between variables measured at the interval and ratio levels. On a ratio scale, 10 is 2 points higher than 8 and is also 2 *times* greater than 5 —the numbers can be compared in a ratio. Ratio numbers can be added and subtracted, and because the numbers begin at an absolute zero point, they can be multiplied and divided (so ratios can be formed between the numbers). For example, people's ages can be represented by values ranging from 0 years (or some fraction of a year) to 120 or more. A person who is 30 years old is 15 years older than someone who is 15 years old (30 – 15 = 15) and is twice as old as that person (30/15 = 2). Of course, the numbers also are mutually exclusive and exhaustive, so that every case can be assigned one and only one value.

Continuous measure: A measure with numbers indicating the values of variables as points on a continuum.

It's tempting to accept the numbers that represent the values of a variable measured at the ratio level at face value, but the precision of the numbers can't make us certain about their accuracy. Income data provided in the U.S. Census are often incomplete (Scott 2001); the unemployment rate doesn't account for people who have given up looking for work (Zitner 1996); and the Consumer Price Index (CPI) does not reflect the types of goods that many groups of consumers buy (Uchitelle 1997). In each of these cases, we have to be sure that the measures that we use reflect adequately the concepts that we intend.

The Special Case of Dichotomies

Dichotomies, variables having only two values, are a special case from the standpoint of levels of measurement. The values or attributes of a variable such as “region of origin” clearly vary in kind or quality but not in amount. Thus, the variable is categorical—measured at the nominal level. Yet we can also think of the variable as indicating the presence of the attribute *northern* (or *southern*) or not. Viewed in this way, there is an inherent order: Someone born in the South has more of the southern attribute (it is present) than a northerner (the attribute is not present). It’s also possible to think of a dichotomy as representing an interval level of measurement because there is an equal interval between the two attributes. So what do you answer to the test question, “What is the level of measurement of *region of origin?*” “Nominal,” of course, but you’ll find that when a statistical procedure requires that variables be quantitative, a dichotomy can be perfectly acceptable. And, of course, if you’re thinking ahead, you already recognize that *region of origin* could be defined in a way that has more than two categories—and then it is not a dichotomy. But that just reinforces the key point of the next section: Level of measurement can be understood only in terms of the attributes or values of the variable.

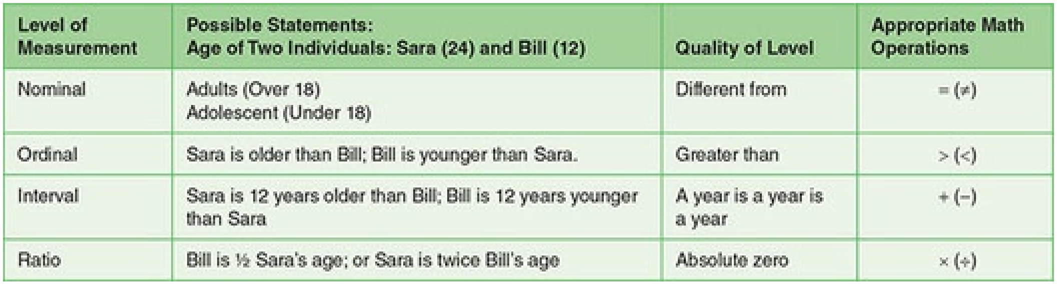
Comparison of Levels of Measurement

Exhibit 4.12 summarizes the types of comparisons that can be made with different levels of measurement, as well as the mathematical operations that are legitimate. Each higher level of measurement allows a more precise mathematical comparison to be made between the value measured at that level compared with those measured at lower levels. However, each comparison between cases measured at lower levels can also be made about cases measured at the higher levels. Thus, all four levels of measurement allow researchers to assign different values to different cases. All three quantitative measures allow researchers to rank cases in order.

Dichotomy: Variable having only two values.

Researchers choose the levels of measurement in the process of operationalizing variables; the level of measurement is not inherent in the variable itself. Many variables can be measured at different levels, with different procedures. For example, the Core Alcohol and Drug Survey (Core Institute 1994) identifies binge drinking by asking students, “Think back over the last two weeks. How many times have you had five or more drinks at a sitting?” You might be ready to classify this as a ratio-level measure, but this would be true only if responses are recorded as the actual number of “times.” Instead, the Core Survey treats this as a closed-ended question, and students are asked to indicate their answer by checking “None,” “Once,” “Twice,” “3 to 5 times,” “6 to 9 times,” or “10 or more times.” Use of these categories makes the level of measurement ordinal, because the distance between any two cases cannot be clearly determined. A student with a response in the “6 to 9 times” category could have binged just one more time than a student who responded “3 to 5 times.” You just can’t tell.

Exhibit 4.12 Properties of Levels of Measurement



Level of Measurement	Possible Statements: Age of Two Individuals: Sara (24) and Bill (12)	Quality of Level	Appropriate Math Operations
Nominal	Adults (Over 18) Adolescent (Under 18)	Different from	= (≠)
Ordinal	Sara is older than Bill; Bill is younger than Sara.	Greater than	> (<)
Interval	Sara is 12 years older than Bill; Bill is 12 years younger than Sara	A year is a year is a year	+ (−)
Ratio	Bill is ½ Sara's age; or Sara is twice Bill's age	Absolute zero	× (÷)

Source: Adapted from material provided by Tajuana D. Massie, assistant professor, social sciences, South Carolina State University.

It is usually a good idea to try to measure variables at the highest level of measurement possible. The more information available, the more ways we have to compare cases. We also have more possibilities for statistical analysis with quantitative than with qualitative variables. Thus, if doing so does not distort the meaning of the concept that is to be measured, measure at the highest level possible. Even if your primary concern is only to compare teenagers with young adults, measure age in years rather than in categories; you can always combine the ages later into categories corresponding to teenager and young adult.

Be aware, however, that other considerations may preclude measurement at a higher level. For example, many people are very reluctant to report their exact incomes, even in anonymous questionnaires. So asking respondents to report their income in categories (such as less than $10,000, $10,000–$19,999, $20,000–$29,999) will result in more responses, and thus more valid data, than will asking respondents for their income in dollars.

Often, researchers treat variables measured at the interval and ratio levels as comparable. They then refer to this as the interval–ratio level of measurement. You will learn in Chapter 14 that different statistical procedures are used for variables with fixed measurement units, but it usually doesn’t matter whether there is an absolute zero point.

Evaluating Measures

Do the operations developed to measure our variables actually do so—are they valid? If we have weighed our measurement options, carefully constructed our questions and observational procedures, and selected sensibly from the available data indicators, we should be on the right track. But we cannot have much confidence in a measure until we have empirically evaluated its validity. What good is our measure if it doesn't measure what we think it does? If our measurement procedure is invalid, we might as well go back to the starting block and try again. As a part of evaluating the validity of our measures, we must also evaluate their reliability, because reliability (consistency) is a prerequisite for measurement validity.

Interval–ratio level of measurement: A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but may not have an absolute, or fixed, zero point.

Measurement Validity

In Chapter 2, you learned that measurement validity refers to the extent to which measures indicate what they are intended to measure. More technically, a valid measure of a concept is one that is closely related to other apparently valid measures of the concept and to the known or supposed correlates of that concept, but that is not related to measures of unrelated concepts, irrespective of the methods used for the other different measures (Brewer and Hunter 1989:134).

When a measure "misses the mark"—when it is not valid—our measurement procedure has been affected by measurement error. Measurement error has two sources:

1. Idiosyncratic (or "random") errors are errors that affect individuals or cases in unique ways that are unlikely to be repeated in just the same way (Viswanathan 2005:289). Individuals make idiosyncratic errors when they don't understand a question, when some unique feelings are triggered by the wording of a question, or when they are feeling out of sorts because of some recent events. Idiosyncratic errors may arise in observational research when the observer is distracted or misperceives an event. In coding content, an idiosyncratic error may be made in recording a numerical code or in skipping a page or web screen. Some idiosyncratic error is unavoidable with any measurement procedure, although it is important to reduce their size as much as possible. However, because such errors are idiosyncratic, or random, they are as likely to be above as below the true value of the measure; this means that idiosyncratic errors should not bias the measure in one direction.
2. Systematic errors occur when responses are affected by factors that are not what the instrument is intended to measure. For example, individuals who like to please others by giving socially desirable responses may have a tendency to say that they "agree" with the statements, simply because they try to avoid saying they "disagree" with anyone. Systematic errors may also arise when the same measure is used across cultures that differ in their understanding of the concepts underlying the measures (Church 2010:152–153). In addition, questions that are unclear may be misinterpreted by most respondents, while unbalanced response choices may lead most respondents to give positive rather than negative responses. For example, if respondents are asked the question with the unbalanced response choices in Exhibit 4.13, they are more likely to respond that gun ownership is wrong than if they are asked the question with the balanced response choices (Viswanathan 2005:142–148).

Systematic errors can do much more damage to measurement validity than idiosyncratic errors can because they will lead to the average value of the indicator being higher or lower than the phenomenon that it is measuring. For example, if a political poll uses questions that generate agreement bias, a politician may believe that voters agree with her position much more than they actually do.

The social scientist must try to reduce measurement errors and then to evaluate the extent to which the resulting measures are valid. The extent to which measurement validity has been achieved can be assessed with four different approaches: (1) face validation, (2) content validation, (3) criterion validation, and (4) construct validation. The methods of criterion and construct validation also include subtypes.

Idiosyncratic (or "random") errors: Errors that affect individuals or other cases in unique ways that are unlikely to be repeated in just the same way.

Systematic errors: Errors that result from factors that are not what an instrument is intended to measure and that affect individuals or other cases in ways that are likely to recur in just the same way, thus creating measurement bias.

Unbalanced response choices: When a fixed-choice survey question has a different number of positive and negative response choices.

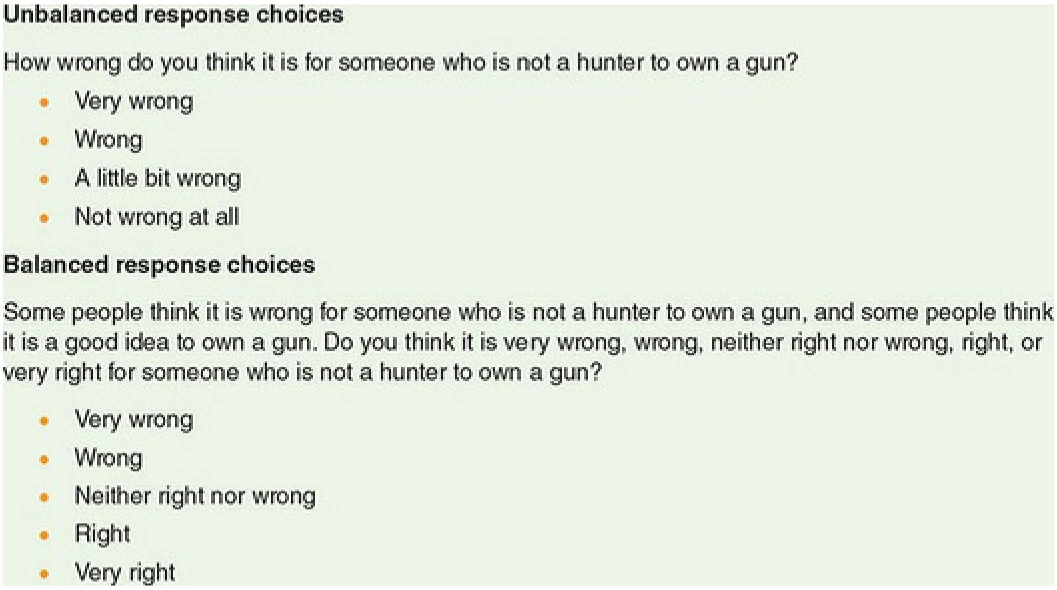
Balanced response choices: When a fixed-choice survey question has an equal number of responses to express positive and negative choices in comparable language.

Face Validity

Researchers apply the term face validity to the confidence gained from careful review of a measure to see if it seems appropriate "on its face." More precisely, we can say that a measure is face valid if it obviously pertains to the meaning of the concept being measured more than to other concepts (Brewer and Hunter 1989:131). For example, a count of the number of drinks people had consumed in the past week would be a face-valid measure of their alcohol consumption. But speaking of "face" validity, what would you think about assessing the competence of political candidates by how mature their faces look? It turns out that people are less likely to vote for candidates with more "baby-faced" features, such as rounded features and large eyes, irrespective of the candidates' records (Cook 2005). It's an unconscious bias, and, of course, it's not one that we would use as a basis for assessing competence in a social science study!

Face validity: The type of validity that exists when an inspection of items used to measure a concept suggests that they are appropriate "on their face."

Exhibit 4.13 Balanced and Unbalanced Response Choices



Unbalanced response choices

How wrong do you think it is for someone who is not a hunter to own a gun?

- Very wrong
- Wrong
- A little bit wrong
- Not wrong at all

Balanced response choices

Some people think it is wrong for someone who is not a hunter to own a gun, and some people think it is a good idea to own a gun. Do you think it is very wrong, wrong, neither right nor wrong, right, or very right for someone who is not a hunter to own a gun?

- Very wrong
- Wrong
- Neither right nor wrong
- Right
- Very right

Although every measure should be inspected in this way, face validation in itself does not provide convincing evidence of measurement validity (DeVellis 2017:100–102). The question "How much beer or wine did you have to drink last week?" looks valid on its face as a measure of frequency of drinking, but people who drink heavily tend to underreport the amount they drink. So the question would be an invalid measure, at least in a study of heavy drinkers. More generally, it is not clear whose impressions are valid "on their face." For example, a doctor and a patient may have very different opinions of the validity of a question about health status.

Content Validity

Content validity establishes that the measure covers the full range of the concept's meaning. To determine that range of meaning, the researcher may solicit the opinions of experts and review literature that identifies the different aspects, or dimensions, of the concept.

An example of a measure that covers a wide range of meaning is the Michigan Alcoholism Screening Test (MAST). The MAST includes 24 questions representing the following subscales: recognition of alcohol problems by self and others; legal, social, and work problems; help seeking; marital and family difficulties; and liver pathology (Skinner and Sheu 1982). Many experts familiar with the direct consequences of substance abuse agree that these dimensions capture the full range of possibilities. Thus, the MAST is believed to be valid from the standpoint of content validity.

Content validity: The type of validity that exists when the full range of a concept's meaning is covered by the measure.

Criterion Validity

When people drink an alcoholic beverage, the alcohol is absorbed into their blood and then gradually metabolized (broken down into other chemicals) in their livers (NIAAA 1997). The alcohol that remains in their blood at any point, unmetabolized, impairs both thinking and behavior (NIAAA 1994). As more alcohol is ingested, cognitive and behavioral consequences multiply. The bases for these biological processes can be identified with direct measures of alcohol concentration in the blood, urine, or breath. Questions about the quantity and frequency of drinking can be viewed as attempts to measure indirectly what biochemical tests measure directly.

Criterion validity is established when the scores obtained on one measure can be compared accurately with those obtained with a more direct or already validated measure of the same phenomenon (the criterion). A measure of blood alcohol concentration or a urine test could serve as the criterion for validating a self-report measure of drinking, as long as the questions we ask about drinking refer to the same time period. Chemical analysis of hair samples can reveal unacknowledged drug use (Mieczkowski 1997). Friends' or relatives' observations of a person's substance use also could serve, in some limited circumstances, as a criterion for validating self-report substance use measures.

Criterion validation studies of self-reported substance abuse measures have yielded inconsistent results. Self-reports of drug use agreed with urinalysis results for about 85% of the drug users who volunteered for a health study in several cities (Weatherby et al. 1994). However, the posttreatment drinking behavior self-reported by 100 male alcoholics was substantially less than the drinking behavior observed by the alcoholics' friends or relatives (Watson et al. 1984). College students' reports of drinking are suspect too: A standard question to measure alcohol use is to ask respondents how many glasses they consume when they do drink. A criterion validation study of this approach measured how much of the drink students poured when they had what they considered to be a "standard" drink (White et al. 2003). The students consistently overestimated how much fluid goes into a standard drink.

Inconsistent findings about the validity of a measure can occur because of differences in the adequacy of a measure across settings and populations. We cannot simply assume that a measure that was validated in one study is also valid in another setting or with a different population. The validity of even established measures has to be tested when they are used in a different context (Viswanathan 2005:297).

The criterion that researchers select can be measured either at the same time as the variable to be validated or after that time. Concurrent validity exists when a measure yields scores that are closely related to scores on a criterion measured at the same time. A store might validate a question-based test of sales ability by administering it to sales personnel who are already employed and then comparing their test scores with their sales performance. Or a measure of walking speed based on mental counting might be validated concurrently with a stopwatch. Predictive validity is the ability of a measure to predict scores on a criterion measured in the future. For example, a store might administer a test of sales ability to new sales personnel and then validate the measure by comparing these test scores with the criterion—the subsequent sales performance of the new personnel.

An attempt at criterion validation is well worth the effort because it greatly increases confidence that the standard is measuring what was intended. However, for many concepts of interest to social scientists, no other variable can reasonably be considered a criterion. If we are measuring feelings or beliefs or other subjective states, such as feelings of loneliness, what *direct* indicator could serve as a criterion? Even with variables for which a reasonable criterion exists, the researcher may not be able to gain access to the criterion—as would be the case with a tax return or employer document that we might wish we could use as a criterion for self-reported income.

Criterion validity: The type of validity that is established by comparing the scores obtained on the measure being validated with those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

Concurrent validity: The type of validity that exists when scores on a measure are closely related to scores on a criterion measured at the same time.

Predictive validity: The type of validity that exists when a measure predicts scores on a criterion measured in the future.

Construct Validity

Measurement validity can also be established by showing that a measure is related to a variety of other measures as specified in a theory. This validation approach, known as construct validity, is commonly used in social research when no clear criterion exists for validation purposes. For example, in one study of the validity of the Addiction Severity Index (ASI), A. Thomas McLellan and his associates (1985) compared subject scores on the ASI with a number of indicators that they felt, from prior research, should be related to substance abuse: medical problems, employment problems, legal problems, family problems, and psychiatric problems. The researchers could not use a criterion validation approach because they did not have a more direct measure of abuse, such as laboratory test scores or observer reports. However, their extensive research on the subject had given them confidence that these sorts of problems were all related to substance abuse, and, indeed, they found that individuals with higher ASI ratings tended to have more problems in each of these areas.

Two other approaches to construct validation are convergent validation and discriminant validation. Convergent validity is achieved when one measure of a concept is associated with different types of measures of the same concept (this relies on the same type of logic as measurement triangulation). Discriminant validity is a complementary approach to construct validation. In this approach, scores on the measure to be validated are compared with scores on measures of different but related concepts. Discriminant validity is achieved if the measure to be validated is not associated strongly with the measures of different concepts. McLellan et al. (1985) found that the ASI passed the tests of convergent and discriminant validity: The ASI's measures of alcohol and drug problems were related more strongly to other measures of alcohol and drug problems than they were to measures of legal problems, family problems, medical problems, and the like.

The distinction between criterion validation and construct validation is not always clear. Opinions can differ about whether a particular indicator is indeed a criterion for the concept that is to be measured. For example, if you need to validate a question-based measure of sales ability for applicants to a sales position, few would object to using actual sales performance as a criterion. But what if you want to validate a question-based measure of the amount of social support that people receive from their friends? Could friends' reports of the amount of support they provided serve as a criterion for validating the amount of support that people say they have received? Are verbal accounts of the amount of support provided adequate? What about observation of social support that people receive? Even if you could observe people in the act of counseling or otherwise supporting their friends, can observers be sure that the interaction is indeed supportive, or that they have observed all the relevant interactions? There isn't really a criterion here, only related concepts that could be used in a construct validation strategy. Even biochemical measures of substance abuse are questionable as criteria for validating self-reported substance use. Urine test results can be altered by ingesting certain substances, and blood tests vary in their sensitivity to the presence of drugs over a particular period.

What both construct validation and criterion validation have in common is the comparison of scores on one measure with the scores on other measures that are predicted to be related. It is not so important that researchers agree that a particular comparison measure is a criterion rather than a related construct. But it is very important to think critically about the quality of the comparison measure and whether it actually represents a different view of the same phenomenon. For example, correspondence between scores on two different self-report measures of alcohol use is a much weaker indicator of measurement validity than is the correspondence of a self-report measure with an observer-based measure of substance use.

Measurement Reliability

Reliability means that a measurement procedure yields consistent scores when the phenomenon being measured is not changing (or that the measured scores change in direct correspondence to actual changes in the phenomenon). If a measure is reliable, it is affected less by random error, or chance variation, than if it is unreliable. Reliability is a prerequisite for measurement validity: We cannot really measure a phenomenon if the measure we are using gives inconsistent results. Because it is usually easier to assess reliability than validity, you are more likely to see an evaluation of measurement reliability in a research report than an evaluation of measurement validity.

Construct validity: The type of validity that is established by showing that a measure is related to other measures as specified in a theory.

Convergent validity: An approach to construct validation; the type of validity achieved when one measure of a concept is associated with different types of measures of the same concept.

Discriminant validity: An approach to construct validation; the scores on the measure to be validated are compared with scores on another measure of the same variable and to scores on variables that measure different but related concepts. Discriminant validity is achieved if the measure to be validated is related most strongly to its comparison measure and less so to the measures of other concepts.

Reliability: When a measurement procedure yields consistent scores when the phenomenon being measured is not changing.

Problems in reliability can occur when inconsistent measurements are obtained after the same phenomenon is measured multiple times, with multiple indicators, or by multiple observers. For example, a test of your knowledge of research methods would be unreliable if every time you took it, you received a different score even though your knowledge of research methods had not changed in the interim, not even as a result of taking the test more than once. This is test–retest reliability. A measure also would be unreliable if slightly different versions of it resulted in markedly different responses (it would not achieve alternate-forms reliability). Similarly, an index composed of questions to measure knowledge of research methods would be unreliable if respondents' answers to each question were totally independent of their answers to the others. By contrast, the index has interitem reliability if the component items are closely related. Finally, an assessment of the level of conflict in social groups would be unreliable if ratings of the level of conflict by two observers were not related to each other (it would then lack interobserver reliability).

Multiple Times: Test–Retest and Alternate Forms

When researchers measure a phenomenon that does not change between two points separated by an interval of time, the degree to which the two measurements are related to each other is the test–retest reliability of the measure. If you take a test of your math ability and then retake the test 2 months later, the test is performing reliably if you receive a similar score both times—presuming that nothing happened during the 2 months to change your math ability. Of course, if events between the test and the retest have changed the variable being measured, then the difference between the test and retest scores should reflect that change.

When ratings by an observer, rather than ratings by the subjects themselves, are being assessed at two or more points in time, test–retest reliability is termed intrarater (or intraobserver) reliability.

If an observer's ratings of individuals' drinking behavior in bars are similar at two or more points in time, and the behavior has not changed, the observer's ratings of drinking behavior are reliable.

One example of how evidence about test–retest reliability may be developed is a study by Linda Sobell and her associates (1988) of alcohol abusers' past drinking behavior (using the Lifetime Drinking History Questionnaire) and life changes (using the Recent Life Changes Questionnaire). All 69 subjects in the study were patients in an addiction treatment program. They had not been drinking before the interview (determined by a breath test). The two questionnaires were administered by different interviewers about 2 or 3 weeks apart, both times asking the subjects to recall events 8 years before the interviews. Reliability was high: 92% of the subjects reported the same life events both times, and at least 81% of the subjects were classified consistently at both interviews as having had an alcohol problem or not. When asked about their inconsistent answers, subjects reported that in the earlier interview they had simply dated an event incorrectly, misunderstood the question, evaluated the importance of an event differently, or forgotten an event. Answers to past drinking questions were less reliable when they were very specific, apparently because the questions exceeded the subjects' capacities to remember accurately.

Researchers test alternate-forms reliability by comparing the subjects' answers with slightly different versions of the survey questions (Litwin 1995:13–21). A researcher may reverse the order of the response choices in an index or modify the question wording in minor ways and then readminister that index to the subjects. If the two sets of responses are not too different, alternate-forms reliability is established.

Test–retest reliability: A measurement showing that measures of a phenomenon at two points in time are highly correlated, if the phenomenon has not changed or has changed only as much as the measures have changed.

Intrarater (or intraobserver) reliability: Consistency of ratings by an observer of an unchanging phenomenon at two or more points in time.

Alternate-forms reliability: A procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

Multiple Indicators: Interitem and Split-Half

When researchers use multiple items to measure a single concept, they must be concerned with interitem reliability (or internal consistency). For example, if we are to have confidence that a set of questions (such as those in Exhibit 4.14) reliably measures depression, the answers to these questions should be highly associated with one another. The stronger the association between the individual items and the more items included, the higher the reliability of the index. Cronbach's alpha is a reliability measure commonly used to measure interitem reliability. Of course, interitem reliability cannot be computed if only one question is used to measure a concept. For this reason, it is much better to use a multi-item index to measure an important concept (Viswanathan 2005:298–299).

Donald Hawkins, Paul Amato, and Valarie King (2007:1007) used the "CES-D" index to measure depression in their study of adolescent well-being and obtained a high level of interitem reliability. They measured "negative outlook" with a similar set of questions (see Exhibit 4.14), but the interitem reliability of this set was lower. Read through the two sets of questions. Do the sets seem to cover what you think of as being depressed and having a negative outlook? If so, they seem to be *content valid* to you.

A related test of reliability is the split-half reliability approach. After a set of survey questions intended to form an index is administered, the researcher divides the questions into half by distinguishing even- and odd-numbered questions, flipping a coin, or using some other random procedure. Scores are then computed for these two sets of questions. The researchers then compare the scores for the two halves and check the relation between the subjects' scores on them. If scores on the two halves are similar and highly related to each other (so that people who score high on one half also score high on the other half, etc.), then the measure's split-half reliability is established.

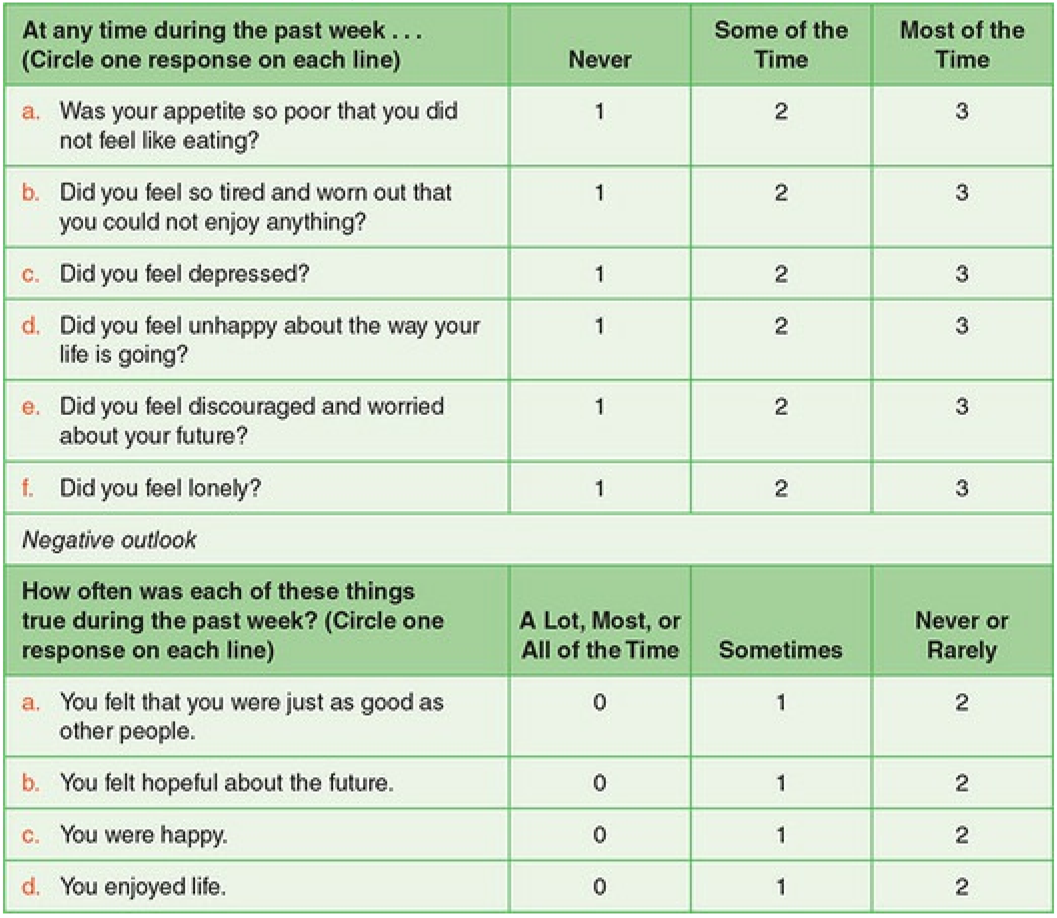
Interitem reliability: An approach that calculates reliability based on the correlation among multiple items used to measure a single concept; also known as internal consistency.

Cronbach's alpha: A statistic commonly used to measure interitem reliability.

Reliability measures: Statistics that summarize the consistency among a set of measures; Cronbach's alpha is the most common measure of the reliability of a set of items included in an index.

Split-half reliability: Reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.

Exhibit 4.14 Examples of Indexes: Short Form of the Center for Epidemiologic Studies (CES-D) and “Negative Outlook” Index



At any time during the past week . . . (Circle one response on each line)	Never	Some of the Time	Most of the Time
a. Was your appetite so poor that you did not feel like eating?	1	2	3
b. Did you feel so tired and worn out that you could not enjoy anything?	1	2	3
c. Did you feel depressed?	1	2	3
d. Did you feel unhappy about the way your life is going?	1	2	3
e. Did you feel discouraged and worried about your future?	1	2	3
f. Did you feel lonely?	1	2	3
Negative outlook			
How often was each of these things true during the past week? (Circle one response on each line)	**A Lot, Most, or All of the Time**	**Sometimes**	**Never or Rarely**
a. You felt that you were just as good as other people.	0	1	2
b. You felt hopeful about the future.	0	1	2
c. You were happy.	0	1	2
d. You enjoyed life.	0	1	2

Source: Hawkins, Daniel N., Paul R. Amato, and Valarie King. 2007. “Nonresident Father Involvement and Adolescent Well-Being: Father Effects or Child Effects?” American Sociological Review 72:990.

Multiple Observers: Interobserver and Intercoder

When researchers use more than one observer to rate the same people, events, or places, interobserver reliability is their goal. If observers are using the same instrument to rate the same thing, their ratings should be very similar. If they are similar, we can have much more confidence that the ratings reflect the phenomenon being assessed rather than the orientations of the observers.

Assessing interobserver reliability is most important when the rating task is complex. Consider a measure of neighborhood disorder, the African American Health Neighborhood Assessment Scale (AAH NAS), which is shown in Exhibit 4.15. The rating task seems straightforward, with clear descriptions of the block characteristics that are rated to produce an overall neighborhood score. However, the judgments that the rater must make while using this scale are complex. They are affected by a wide range of subject characteristics, attitudes, and behaviors as well as by the rater's reactions. As a result, although interobserver agreement on the AAH NAS can be sufficient to achieve a reasonable level of reliability, achieving reliable ratings required careful training of the raters. Qualitative researchers confront the same issue when multiple observers attempt to observe similar phenomena in different settings or at different times. We return to this issue in the qualitative research chapters.

It is also important to establish an adequate level of intercoder reliability when data are transferred from their original form, whether observations or interviews, into structured codes or simply into a data entry program or spreadsheet. There can be weak links in data processing, so the consistency of coders should be tested.

Ways to Improve Reliability and Validity

Whatever the concept measured or the validation method used, no measure is without some error, nor can we expect it to be valid for all times and places. For example, the reliability and validity of self-report measures of substance abuse vary with factors such as whether the respondents are sober or intoxicated at the time of the interview, whether the measure refers to recent or lifetime abuse, and whether the respondents see their responses as affecting their chances of receiving housing, treatment, or some other desired outcome (Babor, Stephens, and Marlatt 1987). In addition, persons with severe mental illness are, in general, less likely to respond accurately (Corse, Hirschinger, and Zanis 1995). We should always be on the lookout for ways in which we can improve the reliability and validity of the measures we use.

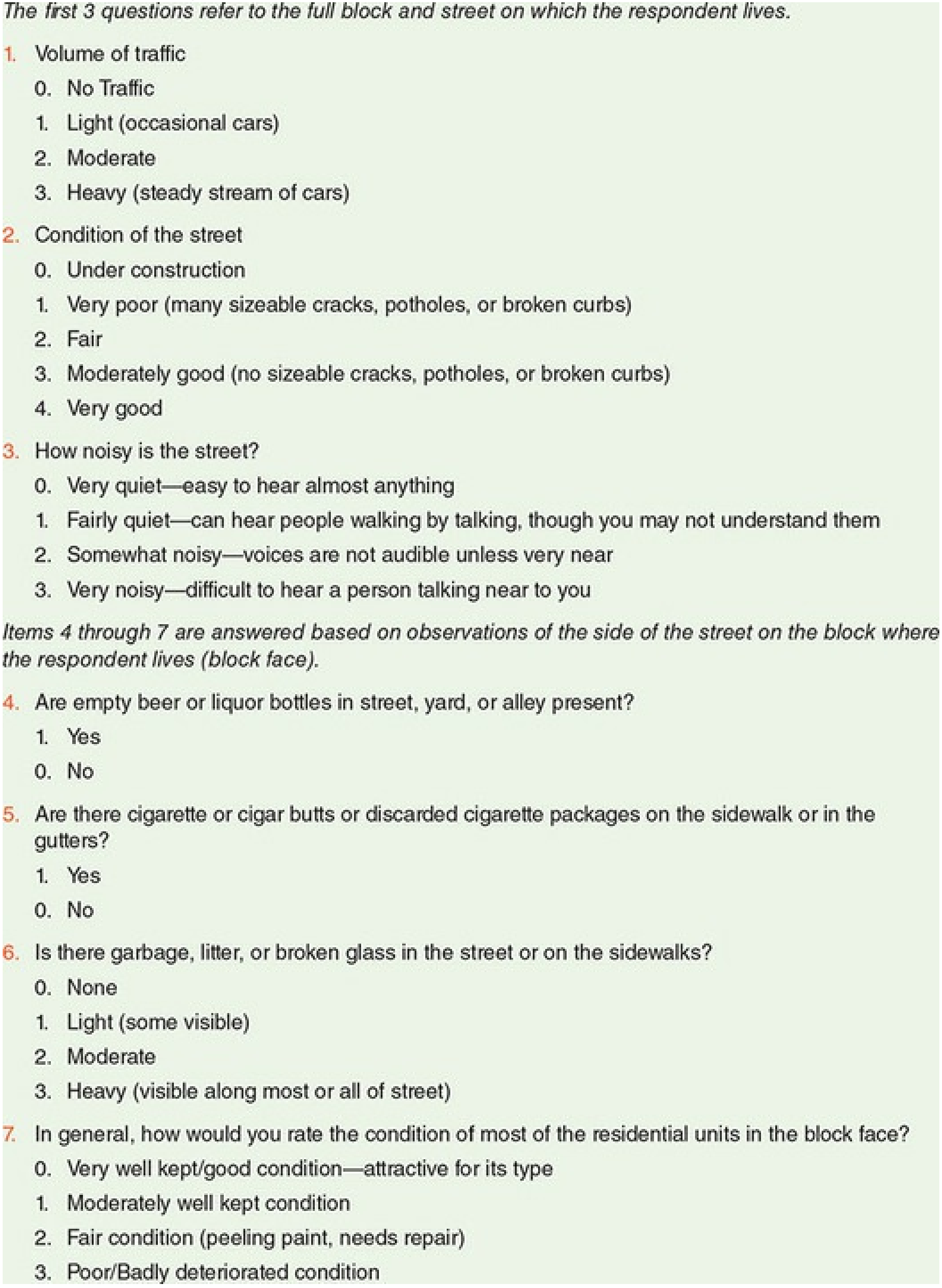
Remember that a reliable measure is not necessarily a valid measure, as Exhibit 4.16 illustrates. This discrepancy is a common flaw of self-report measures of substance abuse. Most respondents answer the multiple questions in self-report indexes of substance abuse in a consistent way, so the indexes are reliable. However, a number of respondents will not admit to drinking, even though they drink a lot. Their answers to the questions are consistent, but they are consistently misleading. As a result, some indexes based on self-report are reliable but invalid. Such indexes are not useful and should be improved or discarded. Unfortunately, many measures are judged to be worthwhile on the basis of only a reliability test.

Interobserver reliability: When similar measurements are obtained by different observers rating the same persons, events, or places.

Intercoder reliability: When the same codes are entered by different coders who are recording the same data.

The reliability and validity of measures in any study must be tested after the fact to assess the quality of the information obtained. But then, if it turns out that a measure cannot be considered reliable and valid, little can be done to save the study. Hence, it is supremely important to select, in the first place, measures that are likely to be reliable and valid. Don't just choose the first measure you find or can think of: Consider the different strengths of different measures and their appropriateness to your study. Conduct a pretest in which you use the measure with a small sample, and check its reliability. Provide careful training to ensure a consistent approach if interviewers or observers will administer the measures. In most cases, however, the best strategy is to use measures that have been used before and whose reliability and validity have been established in other contexts. But the selection of tried and true measures still does not absolve researchers from the responsibility of testing the reliability and validity of the measure in their own studies.

Exhibit 4.15 The Challenge of Interobserver Reliability: African American Health Seven-Item Neighborhood Assessment Scale (AAH NAS)



The first 3 questions refer to the full block and street on which the respondent lives.

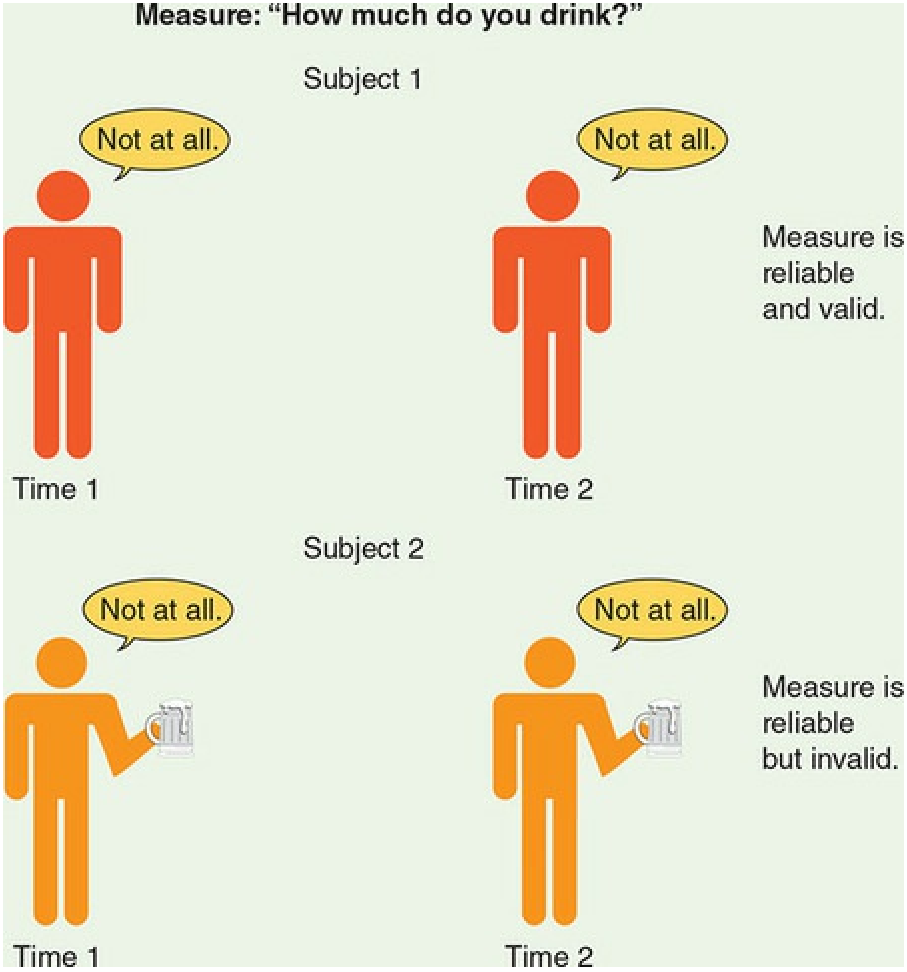
1. Volume of traffic
    0. No Traffic
    1. Light (occasional cars)
    2. Moderate
    3. Heavy (steady stream of cars)
2. Condition of the street
    0. Under construction
    1. Very poor (many sizeable cracks, potholes, or broken curbs)
    2. Fair
    3. Moderately good (no sizeable cracks, potholes, or broken curbs)
    4. Very good
3. How noisy is the street?
    0. Very quiet—easy to hear almost anything
    1. Fairly quiet—can hear people walking by talking, though you may not understand them
    2. Somewhat noisy—voices are not audible unless very near
    3. Very noisy—difficult to hear a person talking near to you

Items 4 through 7 are answered based on observations of the side of the street on the block where the respondent lives (block face).

4. Are empty beer or liquor bottles in street, yard, or alley present?
    1. Yes
    0. No
5. Are there cigarette or cigar butts or discarded cigarette packages on the sidewalk or in the gutters?
    1. Yes
    0. No
6. Is there garbage, litter, or broken glass in the street or on the sidewalks?
    0. None
    1. Light (some visible)
    2. Moderate
    3. Heavy (visible along most or all of street)
7. In general, how would you rate the condition of most of the residential units in the block face?
    0. Very well kept/good condition—attractive for its type
    1. Moderately well kept condition
    2. Fair condition (peeling paint, needs repair)
    3. Poor/Badly deteriorated condition

Source: Andresen E., T. K. Malmstrom, F. D. Wolinsky, M. Schootman, J. P. Miller, and D. K. Miller. “Rating Neighborhoods for Older Adult Health: Results from the African American Health Study.” BMC Public Health 2008, 8:35. doi:10.1186/1471-2458-8-35.

Exhibit 4.16 The Difference Between Reliability and Validity: Drinking Behavior


Measure: "How much do you drink?"
Subject 1
Not at all.
Not at all.
Measure is reliable and valid.
Time 1
Time 2
Subject 2
Not at all.
Not at all.
Measure is reliable but invalid.
Time 1
Time 2

When the population studied or the measurement context differs from that in previous research, instrument reliability and validity may be affected. So the researchers must take pains with the design of their study. For example, test–retest reliability has proved to be better for several standard measures used to assess substance use among homeless persons when the interview was conducted in a protected setting and when the measures focused on factual information and referred to a recent time interval (Drake, McHugo, and Biesanz 1995). Subjects who were younger, female, recently homeless, and less severely afflicted with psychiatric problems were also more likely to give reliable answers.

It may be possible to improve the reliability and validity of measures in a study that already has been conducted if multiple measures were used. For example, in our study of housing for homeless mentally ill persons, funded by the National Institute of Mental Health, we assessed substance abuse with several different sets of direct questions as well as with reports from subjects' case managers and others (Goldfinger et al. 1996). We found that the observational reports were often inconsistent with self-reports and that different self-report measures were not always in agreement; hence, the measures were not valid. A more valid measure was initial reports of lifetime substance abuse problems, which identified all those who subsequently abused substances during the project. We concluded that the lifetime measure was a valid way to identify persons at risk for substance abuse problems. No single measure was adequate to identify substance abusers at a particular point in time during the project. Instead, we constructed a composite of observer and self-report measures that seemed to be a valid indicator of substance abuse over 6-month periods.

If the research focuses on previously unmeasured concepts, new measures will have to be devised. Researchers can use one of three strategies to improve the likelihood that new question-based measures will be reliable and valid (Fowler 1995):

> *Engage potential respondents in group discussions about the questions to be included in the survey.* This strategy allows researchers to check for consistent understanding of terms and to hear the range of events or experiences that people will report.
>
> *Conduct cognitive interviews.* Ask people a test question, then probe with follow-up questions about how they understood the question and what their answer meant.
>
> *Audiotape test interviews during the pretest phase of a survey.* The researchers then review these audiotapes and systematically code them to identify problems in question wording or delivery. (pp. 104–129)

In these ways, qualitative methods help improve the validity of the fixed-response questions used in quantitative surveys.

Conclusions

Remember always that measurement validity is a necessary foundation for social research. Gathering data without careful conceptualization or conscientious efforts to operationalize key concepts often is a wasted effort.

The difficulties of achieving valid measurement vary with the concept being operationalized and the circumstances of the particular study. The examples in this chapter of difficulties in achieving valid measures of substance abuse should sensitize you to the need for caution. However, don't let these difficulties discourage you: Substance abuse is a relatively difficult concept to operationalize because it involves behavior that is socially stigmatized and often illegal. Most other concepts in social research present fewer difficulties. But even substance abuse can be measured adequately with a proper research design.

Planning ahead is the key to achieving valid measurement in your own research; careful evaluation is the key to sound decisions about the validity of measures in others' research. Statistical tests can help determine whether a given measure is valid after data have been collected, but if it appears after the fact that a measure is invalid, little can be done to correct the situation. If you cannot tell how key concepts were operationalized when you read a research report, don't trust the findings. And if a researcher does not indicate the results of tests used to establish the reliability and validity of key measures, remain skeptical.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Alternate-forms reliability 136
Balanced response choices 132
Closed-ended (fixed-choice) question 118
Concept 104
Conceptualization 105
Concurrent validity 134
Constant 112
Construct validity 134
Content validity 133
Continuous measure 129
Convergent validity 135
Criterion validity 134
Cronbach's alpha 137
Dichotomy 130
Discrete measure 128
Discriminant validity 135
Exhaustive 118
Face validity 132
Idiosyncratic (or "random") errors 132
Index 128
Indicator 110
Intercoder reliability 138
Interitem reliability 136
Interobserver reliability 138
Interval level of measurement 128
Interval–ratio level of measurement 131
Intrarater (or intraobserver) reliability 136
Level of measurement 125
Measurement 116
Mutually exclusive 118
Nominal level of measurement 126
Open-ended question 119
Operationalization 109
Ordinal level of measurement 127
Predictive validity 134
Ratio level of measurement 129
Reliability 135
Reliability measures 137
Split-half reliability 137

Systematic errors 132
Test–retest reliability 136
Unbalanced response choices 132
Unobtrusive measure 120

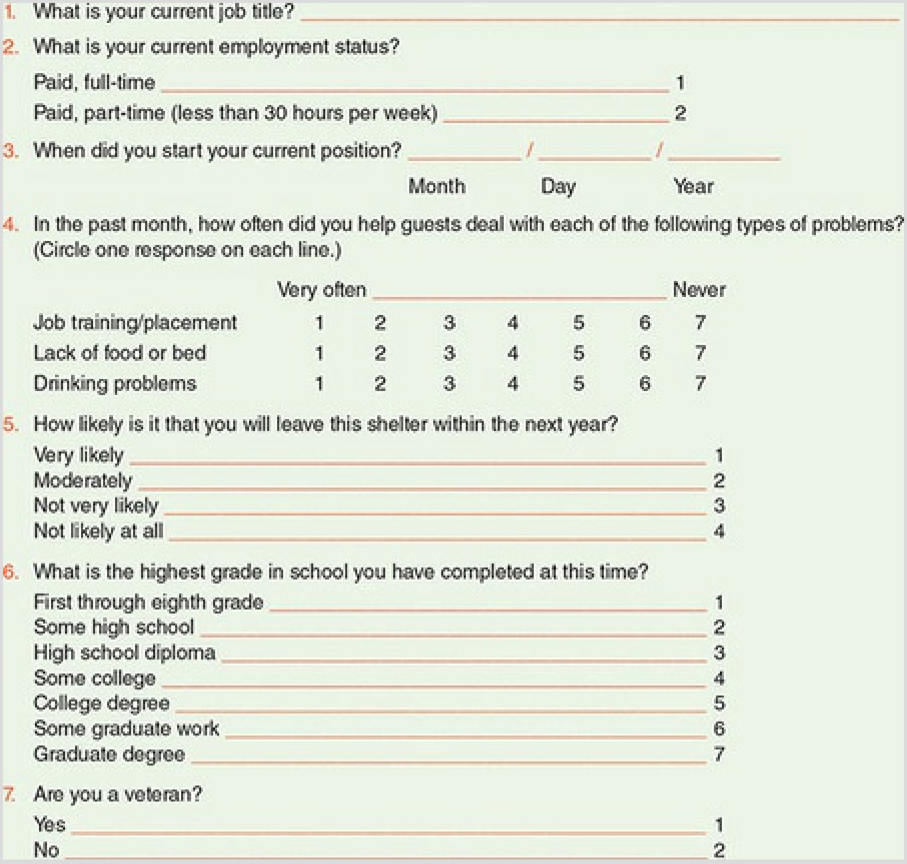
Highlights

- Conceptualization plays a critical role in research. In deductive research, conceptualization guides the operationalization of specific variables; in inductive research, it guides efforts to make sense of related observations.
- Concepts may refer to either constant or variable phenomena. Concepts that refer to variable phenomena may be quite similar to the actual variables used in a study, or they may be much more abstract.
- Concepts are operationalized in research by one or more indicators, or measures, which may derive from observation, self-report, available records or statistics, books and other written documents, clinical indicators, pictures, discarded materials, or some combination of these.
- Single-question measures may be closed-ended, with fixed-response choices, or open-ended, with fixed-response choices and an option to write another response.
- Indexes and scales measure a concept by combining answers to several questions and thus reducing idiosyncratic error variation. Several issues should be explored with every intended index: Does each question actually measure the same concept? Does combining items in an index obscure important relationships between individual questions and other variables? Is the index multidimensional?
- Level of measurement indicates the type of information obtained about a variable and the type of statistics that can be used to describe its variation. The four levels of measurement can be ordered by the complexity of the mathematical operations they permit: nominal (least complex), ordinal, interval, ratio (most complex). The measurement level of a variable is determined by how the variable is operationalized. Dichotomies, a special case, may be treated as measured at the nominal, ordinal, or interval level.
- The validity of measures should always be tested. There are four basic approaches: (1) face validation, (2) content validation, (3) criterion validation (either predictive or concurrent), and (4) construct validation (convergent or discriminant validity). Criterion validation provides the strongest evidence of measurement validity, but there often is no criterion to use in validating social science measures.
- Measurement reliability is a prerequisite for measurement validity, although reliable measures are not necessarily valid. Reliability can be assessed through a test–retest procedure, by interitem consistency, through a comparison of responses to alternate forms of the test, or by consistency among observers.

Discussion Questions

1. What does poverty mean to you? Have you been using an absolute, relative, or objective standard when you think about poverty? Identify two examples of “poverty” and explain why they represent this concept. Compare your conceptualization with those of your classmates and what you find in a dictionary. Can you improve your conceptualization based on some feedback?
2. What questions would you ask to measure feelings of being “in” or “out” with regard to a group? Write five questions for an index and suggest response choices for each. How would you validate this measure using a construct validation approach? Can you think of a criterion validation procedure for your measure?
3. If you were given a questionnaire right now that asked you about your use of alcohol and illicit drugs in the past year, would you disclose the details fully? How do you think others would respond? What if the questionnaire was anonymous? What if there was a confidential ID number on the questionnaire so that the researcher could keep track of who responded? What criterion validation procedure would you suggest for assessing measurement validity?

Exhibit 4.17 Selected Shelter Staff Survey Questions



1. What is your current job title? ____________________

2. What is your current employment status?

Paid, full-time ____________________ 1
Paid, part-time (less than 30 hours per week) ____________________ 2

3. When did you start your current position? ________ / ________ / ________
Month Day Year

4. In the past month, how often did you help guests deal with each of the following types of problems? (Circle one response on each line.)

	Very often						Never
Job training/placement	1	2	3	4	5	6	7
Lack of food or bed	1	2	3	4	5	6	7
Drinking problems	1	2	3	4	5	6	7

5. How likely is it that you will leave this shelter within the next year?

Very likely ____________________ 1
Moderately ____________________ 2
Not very likely ____________________ 3
Not likely at all ____________________ 4

6. What is the highest grade in school you have completed at this time?

First through eighth grade ____________________ 1
Some high school ____________________ 2
High school diploma ____________________ 3
Some college ____________________ 4
College degree ____________________ 5
Some graduate work ____________________ 6
Graduate degree ____________________ 7

7. Are you a veteran?

Yes ____________________ 1
No ____________________ 2

Source: Based on Schutt 1992: 7–10, 15, 16. Results reported in Schutt, R. K. and M. L. Fennell. 1992. “Shelter Staff Satisfaction With Services, the Service Network and Their Jobs.” Current Research on Occupations and Professions 7:177–200.

4. The questions in Exhibit 4.17 are selected from my survey of shelter staff (Schutt and Fennell 1992). First, identify the level of measurement for each question. Then, rewrite each question so that it measures the same variable but at a different level. For example, you might change the question that measures seniority at the ratio level (in years, months, and days) to one that measures seniority at the ordinal level

(in categories). Or you might change a variable measured at the ordinal level, such as highest grade in school completed, to one measured at the ratio level. For the variables measured at the nominal level, try to identify at least two underlying quantitative dimensions of variation, and write questions to measure variation along these dimensions. For example, you might change the question asking, “What is your current job title?” to two questions that ask about the pay in the respondent’s current job and the extent to which the job is satisfying.

What are the advantages and disadvantages of phrasing each question at one level of measurement rather than another? Do you see any limitations on the types of questions for which levels of measurement can be changed?

Practice Exercises

1. Now it's time to try your hand at operationalization with survey-based measures. Formulate a few fixed-choice questions to measure variables pertaining to the concepts you researched for the discussion questions, such feeling poor or perceptions of the level of substance abuse in your community. Arrange to interview one or two other students with the questions you have developed. Ask one fixed-choice question at a time, record your interviewee's answer, and then probe for additional comments and clarifications. Your goal is to discover how respondents understand the meaning of the concept you used in the question and what additional issues shape their response to it.

   When you have finished the interviews, analyze your experience: Did the interviewees interpret the fixed-choice questions and response choices as you intended? Did you learn more about the concepts you were working on? Should your conceptual definition be refined? Should the questions be rewritten, or would more fixed-choice questions be necessary to capture adequately the variation among respondents?
2. Now, try index construction. You might begin with some of the questions you wrote for Practice Exercise 1. Try to write about four or five fixed-choice questions that each measure the same concept. Write each question so that it has the same response choices. Now, conduct a literature search to identify an index that another researcher used to measure your concept or a similar concept. Compare your index to the published index. Which seems preferable to you? Why?
3. Develop a plan for evaluating the validity of a measure. Your instructor will give you a copy of a questionnaire actually used in a study. Pick one question, and define the concept that you believe it is intended to measure. Then develop a construct validation strategy involving other measures in the questionnaire that you think should be related to the question of interest—if it measures what you think it measures.
4. What are some of the research questions you could attempt to answer with the available statistical data? Check out the U.S. Census Bureau website (www.census.gov) and then review the Surveys/Programs descriptions. List five questions you could explore with data from the census or one of its surveys. Identify six variables implied by these research questions that you could operationalize with the available data. What are the three factors that might influence variation in these measures, other than the phenomenon of interest? (Hint: Consider how the data are collected.)
5. One quick and easy way to check your understanding of the levels of measurement, reliability, and validity is with the interactive exercises on the study site. First, select one of the "Levels of Measurement" options from the "Interactive Exercises" link on the main menu, and then read the review information at the start of the lesson. You will then be presented with about 10 variables and response choices and asked to identify the level of measurement for each one. If you make a mistake, the program will give a brief explanation about the level of measurement. After you have reviewed one to four of these lessons, repeat the process with one or more of the "Valid and Reliable Measures" lessons.
6. Go to the book's study site and review the Methods section of two of the research articles that you find at edge.sagepub.com/schutt9e. Write a short summary of the concepts and measures used in these studies. Which article provides clearer definitions of the major concepts? Does either article discuss possible weaknesses in measurement procedures?

Ethics Questions

1. The ethical guidelines for social research require that subjects give their *informed consent* before participating in an interview. How "informed" do you think subjects have to be? If you are interviewing people to learn about substance abuse and its impact on other aspects of health, is it okay to just tell respondents in advance that you are conducting a study of health issues? What if you plan to inquire about victimization experiences? Explain your reasoning.
2. Some Homeland Security practices as well as inadvertent releases of web searching records have raised new concerns about the use of unobtrusive measures of behavior and attitudes. If all identifying information is removed, do you think social scientists should be able to study the extent of prostitution in different cities by analyzing police records? How about how much alcohol different types of people use by linking deidentified credit card records to store purchases?

Web Exercises

1. How would you define *alcoholism?* Write a brief definition. Based on this conceptualization, describe a method of measurement that would be valid for a study of alcoholism (alcoholism as you define it). Now go to the website of the National Institute on Alcohol Abuse and Alcoholism and read their definition of an alcohol use disorder (https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders). Is this definition consistent with your definition?

   What are the "facts" about alcoholism presented by the National Council on Alcohol and Drug Dependence (NCADD) at www.ncadd.org? How is alcoholism conceptualized? Based on this conceptualizing, give an example of one method that would be a valid measurement in a study of alcoholism.

   Now look at some of the other related links accessible from the NCADD website. What are some of the different conceptualizations of alcoholism that you find? How does the chosen conceptualization affect one's choice of methods of measurement?
2. What are the latest findings about student substance abuse from the Harvard School of Public Health? Check out http://archive.sph.harvard.edu/cas/AllIndex.html and write a brief report.
3. A list of different measures of substance abuse is available at a site maintained by the National Institute on Alcoholism and Alcohol Abuse, www.niaaa.nih.gov/research/guidelines-and-resources/recommended-alcohol-questions. There is a lengthy discussion of the various self-report instruments for alcohol problem screening among adults at http://pubs.niaaa.nih.gov/publications/AssessingAlcohol/selfreport.htm (Connors and Volk 2004). Read the Connors and Volk article, and pick two of the instruments they discuss (Connors and Volk 2004:27–32). What concept of substance abuse is reflected in each measure? Is either measure multidimensional? What do you think the relative advantages of each measure might be? What evidence is provided about their reliability and validity? What other test of validity would you suggest?

Video Interview Questions

Listen to the researcher interview for Chapter 4 at edge.sagepub.com/schutt9e.

1. What problems does Dana Hunt identify with questions designed to measure frequency of substance abuse and "aggressive feelings"?
2. What could be done to overcome these problems?

SPSS Exercises

1. View the variable information for the variables AGE, CHILDS, PARTYID3, SOCBAR, RACE, and INCOME06 in the GSS2016 file you are using. Click on the "variable list" icon or choose Utilities/Variables from the menu. Choose PARTYID, then SOCBAR. At which levels (nominal/categorical, ordinal, interval, ratio) are each of these variables measured? (By the way, DK means "Don't Know," NA means "No Answer," and NAP means "Not Applicable.")
2. Review the actual questions used to measure four of the variables in Question 1 or in your hypotheses in Chapter 2's SPSS exercise (Question 3). You can find most GSS questions at the Data Explorer website for the General Social Survey (gssdataexplorer.norc.org). Name the variable that you believe each question measures (search under "keyword"). Discuss the face validity and content validity of each question as a measure of its corresponding variable. Explain why you conclude that each measure is valid or not.
3. CONGOV is part of an index involving the following question: How much confidence do you have in
   a. Executive branch of the federal government
   b. U.S. Supreme Court
   c. Congress

   Now answer the following questions:
   a. What is the concept being measured by this index?
   b. Do you agree that each of these variables belongs in the index? Explain.
   c. What additional variables would you like to see included in this index?

Developing a Research Proposal

At this point, you can begin the process of conceptualization and operationalization. You'll need to assume that your primary research method will be conducting a survey. These next steps correspond to Exhibit 3.10, #7.

1. List at least 10 variables that will be measured in your research. No more than two of these should be sociodemographic indicators such as race or age. The inclusion of each variable should be justified in terms of theory or prior research that suggests it would be an appropriate independent or dependent variable or will have some relation to either of these.
2. Write a conceptual definition for each variable. Whenever possible, this definition should come from the existing literature—either a book you have read for a course or the research literature that you have been searching. Ask two class members for feedback on your definitions.
3. Develop measurement procedures for each variable. Several measures should be single questions and indexes that were used in prior research (search the web and the journal literature in Sociological Abstracts or PsycINFO, the online database of Psychological Abstracts [or its full text version, PsycARTICLES]). Make up a few questions and one index yourself. Ask your classmates to answer these questions and give you feedback on their clarity.
4. Propose tests of reliability and validity for four of the measures.

Chapter 5 Sampling and Generalizability

- Research That Matters, Questions That Count
- Sample Planning
  - The Purpose of Sampling
  - Define Sample Components and the Population
  - Evaluate Generalizability
- Research in the News: What Are Best Practices for Sampling Vulnerable Populations?
- Assess the Diversity of the Population
- Consider a Census
- Sampling Methods
  - Probability Sampling Methods
    - *Simple Random Sampling*
    - *Systematic Random Sampling*
    - *Stratified Random Sampling*
    - *Multistage Cluster Sampling*
    - *Probability Sampling Methods Compared*
  - Nonprobability Sampling Methods
    - *Availability (Convenience) Sampling*
- Careers and Research
  - *Quota Sampling*
  - *Purposive Sampling*
  - *Snowball Sampling*
  - Lessons About Sample Quality
  - Generalizability in Qualitative Research
- Sampling Distributions
  - Estimating Sampling Error
  - Sample Size Considerations
- Conclusions

A common technique in journalism is to put a "human face" on a story. For instance, a *New York Times* reporter (Stewart 2017) interviewed Abdul Hasan for a story about New York City's expansion of services for their growing numbers of homeless persons. The reporter found Mr. Hasan in a new drop-in center in Queens, where he gets three meals a day in spite of living on the streets. He said he became homeless after a dispute with his family over sexuality and drinking, and then quit his job as a restaurant cashier when he was unable to maintain his hygiene. He is now receiving assistance from social workers at the center and says that "it's way better than the street."

Research That Matters, Questions That Count

Young adults who are homeless face a number of barriers to gaining housing and maintaining community connections. Lack of a job is one of the greatest such barriers, but little is known about how many homeless young adults are able to obtain a job and what distinguishes them from others who remain unemployed. Kristin Ferguson at the University of California, San Diego, and social scientists at other universities, Kimberly Bender, Sanna Thompson, Elaine Maccio, and David Pollio (2012) designed a research project to investigate this issue.

Ferguson and her collaborators (2012:389–390) decided to interview homeless young adults in five U.S. cities in different regions. The researchers secured the cooperation of multiservice, nonprofit organizations that provide comprehensive services to homeless youth and then approached youth in these agencies and on the streets—accompanied by agency outreach staff. The researchers first told potential participants about the project and then asked whether they were 18 to 24 years old and had been away from home for at least 2 weeks in the previous month. The potential participants who indicated they were interested and eligible for the study were then offered the opportunity to sign a consent form.

One of the survey findings was that young adults in three of the five cities differed in their employment status and sources of income generation. For example, homeless young adults in Los Angeles were more likely to be employed than others were, and Austin young adults were significantly more likely to receive their income from panhandling (Ferguson et al. 2012:400).

1. What other research questions would interest you in a study of this population?
2. How confident are you that what the researchers learned about the (approximately 50) youth surveyed in each of the five cities represents all homeless youth in these cities? Explain your reasoning.
3. The five cities were Los Angeles, Austin, Denver, New Orleans, and St. Louis. Would you consider homeless youth in these cities to be similar to those in other cities? What about in rural areas? In other countries? Explain.

In this chapter, you will learn about procedures for selecting samples from larger populations. By the end of the chapter, you will understand why drawing a representative sample is important but often very difficult, particularly in studies of hard-to-reach groups such as homeless youth. You will also learn strategies for identifying people to interview intensively when representing a larger population is not the key goal. After you finish the chapter, test yourself by reading the 2012 *Youth & Society* article by Kristin Ferguson and her colleagues at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 5 at edge.sagepub.com/schutt9e.

Ferguson, Kristin M., Kimberly Bender, Sanna J. Thompson, Elaine M. Maccio, and David Pollio. 2012. “Employment Status and Income Generation Among Homeless Young Adults: Results From a Five-City, Mixed-Methods Study.” *Youth & Society* 44:385–407.

It is a sad story that at least includes a sense of progress in dealing with the problem and some hope for the future. Several other stories like Mr. Hasan’s and comments by center staff and civic leaders generate a strong sense of the center’s value (see this chapter’s “Research in the News”). However, we don’t know whether the program participants interviewed for the story are like most program participants, most homeless persons in New York, or most homeless persons throughout the United States—or whether they are just several people who caught the eye of this one reporter. In other words, we don’t know how generalizable their stories are, and if we don’t have confidence in generalizability, then the validity of this account of how the program participants became homeless is suspect. Because we don’t know whether their situation is widely shared or unique, we cannot really judge what the account tells us about the social world.

In this chapter, you will learn about sampling methods, the procedures that primarily determine the generalizability of research findings. I first review the rationale for using sampling in social research and consider two circumstances when sampling is not necessary. I then turn to specific sampling methods and when they are most appropriate, using examples from research on homelessness. That discussion is followed by a discussion of sampling distributions, which introduces you to the logic of statistical inference—that is, how to determine the likelihood that sample statistics represent the population from which the sample was drawn. By the chapter's end, you should understand which questions you need to ask to evaluate the generalizability of a study as well as what choices you need to make when designing a sampling strategy. You should also realize that it is just as important to select the "right" people or objects to study as it is to ask participants the right questions.

Sample Planning

You have encountered the problem of generalizability in each of the studies you have read about in this book. For example, Roger Patulny and Claire Seaman (2017) discussed their findings in the ABS-GSS as though they could be generalized to the entire adult population of Australia; Bohyun Joy Jang, Megan Patrick, and Megan Schuler (2017) generalized their findings about substance use and family formation from the Monitoring the Future survey to the entire young adult population in the United States; and Stanley Milgram's (1963) findings about obedience to authority were generalized to the entire world. Whether you are designing a sampling strategy or evaluating someone else's findings, you have to understand how and why researchers decide to sample and what the consequences of these decisions are for the generalizability of the study's findings.

The Purpose of Sampling

Have you ever met, or seen, a homeless person like Abdul Hasan? Perhaps you have encountered many, or know some people who have been homeless. Did you ever wonder if other homeless persons are like those you have encountered yourself? Have you found yourself drawing conclusions about persons who are homeless based on those you have met? Just like the reporter, you know that you shouldn't conclude that all homeless persons are like those you have encountered, but you also know that you can't hope to learn about every homeless person, even in your own city or town.

The purpose of sampling is to generate a set of individuals or other entities that give you a valid picture of all such individuals, or other entities. That is, a sample is a subset of the larger set of individuals or other entities in which you are interested. If you have done a good job of sampling, you will be able to generalize what you have learned from the subset to the larger set from which it was selected.

As researchers, we call the set of individuals or other entities to which we want to generalize our findings the population. For example, a city government may want to describe the city's entire adult homeless population. If, as is usually the case, the government does not have the time or resources to survey all homeless individuals in the city, it may fund a survey of a subset of these individuals. This subset of the population of interest is a sample. The individual members of this sample are called elements, or elementary units.

Define Sample Components and the Population

In many studies, we sample directly from the elements in the population of interest. We may survey a sample of the entire population of students in a school, based on a list obtained from the registrar's office. This list, from which the elements of the population are selected, is termed the sampling frame. The students who are selected and interviewed from that list are the elements.

Population: The entire set of individuals or other entities to which study findings are to be generalized.

Sample: A subset of a population that is used to study the population as a whole.

Elements: The individual members of the population whose characteristics are to be measured.

Sampling frame: A list of all elements or other units containing the elements in a population.

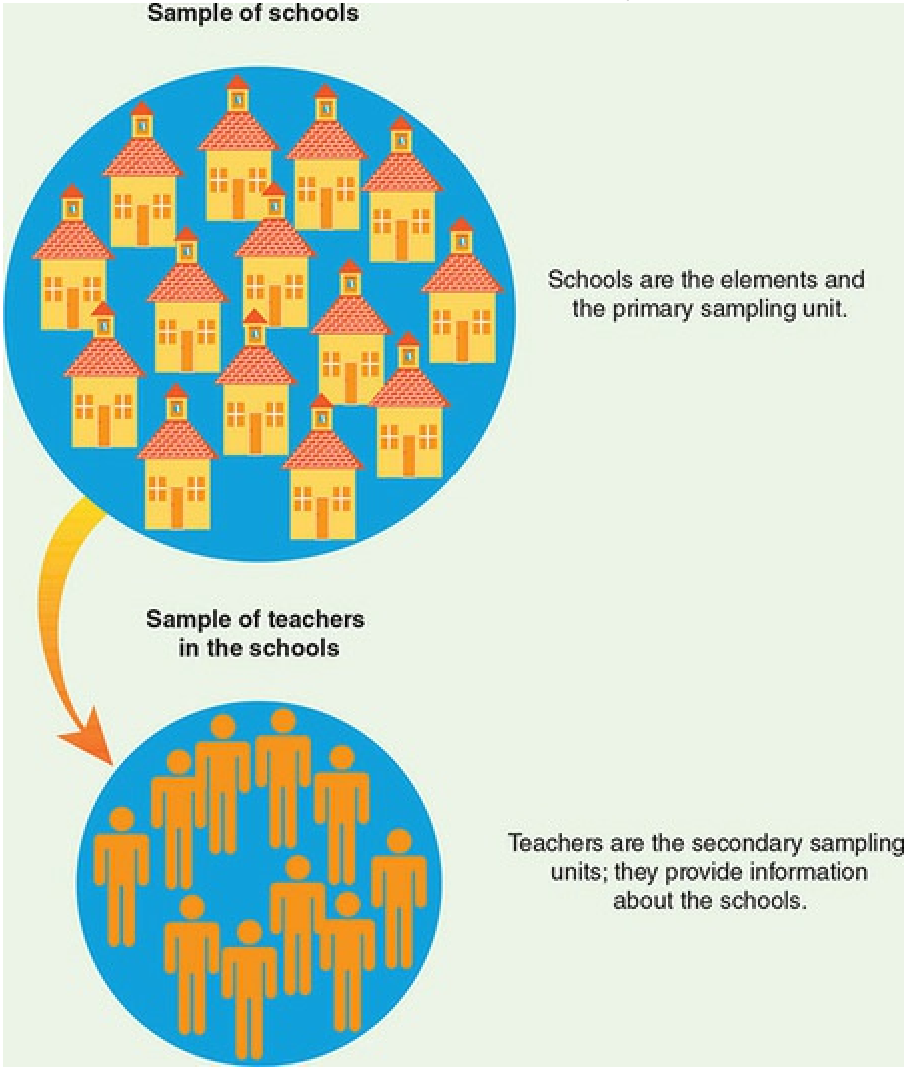
In some studies, the entities that can be reached easily are not the same as the elements from which we want information, but they include those elements. For example, you may have a list of households but not a list of the entire population of a town, even though the adults are the elements that you want to sample. In this situation, you could draw a sample of households so that you can identify the adult individuals in these households. The households are termed enumeration units, and the adults in the households are the elements (Levy and Lemeshow 1999:13–14).

Sometimes, the individuals or other entities from which we collect information are not actually the elements in our study. For example, a researcher might sample schools for a survey about educational practices and then interview a sample of teachers in each sampled school to obtain the information about educational practices. Both the schools and the teachers are termed sampling units, because the researcher samples from both (Levy and Lemeshow 1999:22). The schools are selected in the first stage of the sample, so they are the *primary sampling units* (in this case, they are also the elements in the study). The teachers are *secondary sampling units* (but they are not elements because they are used to provide information about the entire school) (see Exhibit 5.1).

Enumeration units: Units that contain one or more elements and that are listed in a sampling frame.

Sampling units: Units listed at each stage of a multistage sampling design.

Exhibit 5.1 Sample Components in a Two-Stage Study


Sample of schools
Schools are the elements and the primary sampling unit.
Sample of teachers in the schools
Teachers are the secondary sampling units; they provide information about the schools.

Source: Based on information from Levy and Lemeshow (1999).

It is important to know exactly what population a sample can represent when you select or evaluate sample components. In a survey of “adult Americans,” the general population may reasonably be construed as all residents of the United States who are at least 21 years old. But always be alert to ways in which the population may have been narrowed by the sample selection procedures. For example, perhaps only English-speaking residents of the United States were surveyed. The population for a study is the aggregation of elements that we actually focus on and sample from, not some larger aggregation that we really wish we could have studied.

Some populations, such as the homeless, are not identified by a simple criterion such as a geographic boundary or an organizational membership. Clear definition of such a population is difficult but quite necessary. Anyone should be able to determine just what population was actually studied. However, studies of homeless persons in the early 1980s "did not propose definitions, did not use screening questions to be sure that the people they interviewed were indeed homeless, and did not make major efforts to cover the universe of homeless people" (Burt 1996:15). (Perhaps just homeless persons in one shelter were studied.) The result was a "collection of studies that could not be compared" (Burt 1996:15). Several studies of homeless persons in urban areas addressed the problem by employing a more explicit definition of the population: "people who had no home or permanent place to stay of their own (meaning they rented or owned it themselves) and no regular arrangement to stay at someone else's place" (Burt 1996:18).

Even this more explicit definition still leaves some questions unanswered: What is a "regular arrangement"? How permanent does a "permanent place" have to be? In a study of homeless persons in Chicago, Michael Sosin, Paul Colson, and Susan Grossman (1988) answered these questions in their definition of the population of interest:

> We define the homeless as: those current[ly] residing for at least one day but for less than fourteen with a friend or relative, not paying rent, and not sure that the length of stay will surpass fourteen days; those currently residing in a shelter, whether overnight or transitional; those currently without normal, acceptable shelter arrangements and thus sleeping on the street, in doorways, in abandoned buildings, in cars, in subway or bus stations, in alleys, and so forth; those residing in a treatment center for the indigent who have lived at the facility for less than 90 days and who claim that they have no place to go, when released. (p. 22)

This definition reflects accurately Sosin et al.'s concept of homelessness and allows researchers in other locations or at other times to develop procedures for studying a comparable population. The more complete and explicit the definition is of the population from which a sample was selected, the more precise our generalizations can be.

Evaluate Generalizability

Once we have defined clearly the population from which we will sample, we need to determine the scope of the generalizations we will make from our sample. Do you recall from Chapter 2 the two different meanings of generalizability?

Can the findings from a sample of the population be generalized to the population from which the sample was selected?

Did Miller McPherson, Lynn Smith-Lovin, and Matthew Brashears' (2006) findings about social ties apply to the United States, Rich Ling and Gitte Stald's (2010) to all of Norway and Denmark, or Henry Wechsler and colleagues' (2002) study of binge drinking to all U.S. college students? This type of generalizability was defined as *sample generalizability* in Chapter 2.

Can the findings from a study of one population be generalized to another, somewhat different population?

Are mobile phone users in Norway and Denmark similar to those in other Scandinavian countries? In other European countries? Throughout the world? Are students similar to full-time employees, housewives, or other groups in their drinking patterns? Do findings from a laboratory study about obedience to authority at an elite northeastern U.S. college in the 1960s differ from those that would be obtained today at a commuter college in the Midwest? What is the generalizability of the results from a survey of homeless persons in one city? This type of generalizability question was defined as *cross-population generalizability* in Chapter 2.

This chapter focuses attention primarily on the problem of sample generalizability: Can findings from a sample be generalized to the population from which the sample was drawn? This is really the most basic question to ask about a sample, and social research methods provide many tools with which to address it.

Sample generalizability depends on sample quality, which is determined by the amount of sampling error—the difference between the characteristics of a sample and the characteristics of the population from which it was selected. The larger the sampling error, the less representative the sample—and thus the less generalizable the findings. To assess sample quality when you are planning or evaluating a study, ask yourself these questions:

- From what population were the cases selected?
- What method was used to select cases from this population?
- Do the cases that were studied represent, in the aggregate, the population from which they were selected?

Cross-population generalizability (also called “external validity”) involves quite different considerations. Researchers are engaging in cross-population generalizability when they project their findings onto different groups or populations than those they have actually studied. The population to which generalizations are made in this way can be termed the target population—a set of elements different from the population that was sampled and to which the researcher would like to generalize any study findings. When we generalize findings to target populations, we must be somewhat speculative. We must carefully consider the validity of claims that the findings can be applied to other groups, geographic areas, cultures, or times.

The validity of cross-population generalizations cannot be tested empirically, except by conducting more research in other settings, but it is often a concern in qualitative studies and experimental research that does not attempt to sample from a defined population. I discuss the problem of cross-population generalizability later in this chapter, in the context of nonprobability sampling methods that are often used in qualitative research, and then again in Chapter 7, which addresses experimental research, in Chapter 10, which addresses qualitative methods, and in Chapter 15, which discusses methods for studying different societies.

Sampling error: Any difference between the characteristics of a sample and the characteristics of a population; the larger the sampling error, the less representative the sample.

Target population: A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.

Assess the Diversity of the Population

Sampling is unnecessary if all the units in the population are identical. Physicists don’t need to select a representative sample of atomic particles to learn about basic physical processes. They can study a single atomic particle because it is identical to every other particle of its type. Similarly, biologists don’t need to sample a particular type of plant to determine whether a given chemical has toxic effects on that particular type. The idea is “If you’ve seen one, you’ve seen ’em all.”

In the News Research in the News: What are Best Practices for Sampling Vulnerable Populations?

For Further Thought?

A New York City survey estimated 3,900 people living on the street, and the city's Department of Homeless Services (DHS) is opening new drop-in centers to help meet their basic needs. Finding housing prices impossibly high for meager incomes—even for some who are working—street-dwelling homeless persons have often tried and rejected the option of staying in shelters due to experiences with or fear of crime, overcrowding, or other problems. The DHS estimates that it takes an average of 5 months of contact to reestablish trust and convince people to return to living indoors. Although the city is also opening more shelters, some are designated as Safe Havens of limited size in order to attract more of the street homeless.

1. What research question would be of most interest to you that might be the focus of a survey of a sample of homeless persons dwelling on the street?
2. How many challenges can you list that would likely be confronted by a social researcher seeking to survey a representative sample of homeless persons?
3. Can you identify strategies discussed in this chapter for overcoming some of these challenges?

News source: Stewart, Nikita. 2017. "As More Opt for Streets, City Offers a Place to Go." *The New York Times,* July 19, p. A20.

What about people? Certainly, all people are not identical (nor are other animals, in many respects). Nonetheless, if we are studying physical or psychological processes that are the same among all people, sampling is not needed to achieve generalizable findings. Psychologists and social psychologists often conduct experiments on college students to learn about processes that they think are identical across individuals. They believe that most people would have the same reactions as the college students if they experienced the same experimental conditions. Field researchers who observe group processes in a small community sometimes make the same assumption.

There is a potential problem with this assumption, however: There's no way to know for sure if the processes being studied are identical across all people. In fact, experiments can give different results depending on the type of people who are studied or the conditions for the experiment. Milgram's (1965) classic experiments on obedience to authority, which you studied in Chapter 3, illustrate this point very well. You remember that the original

Milgram experiments tested the willingness of male volunteers in New Haven, Connecticut, to comply with the instructions of an authority figure to give "electric shocks" to someone else, even when these shocks seemed to harm the person receiving them. In most cases, the volunteers complied. Milgram concluded that people are very obedient to authority.

Were these results generalizable to all men, to men in the United States, or to men in New Haven? The initial experiment was repeated many times to assess the generalizability of the findings. Similar results were obtained in many replications of the Milgram experiments—that is, when the experimental conditions and subjects were similar to those Milgram studied. Other studies showed that some groups were less likely to react so obediently. Given certain conditions, such as another "subject" in the room who refused to administer the shocks, subjects were likely to resist authority.

So, what do the initial experimental results tell us about how people will react to an authoritarian movement in the real world, when conditions are not so carefully controlled? In the real social world, people may be less likely to react obediently as well. Other individuals may argue against obedience to a particular leader's commands, or people may see on TV the consequences of their actions. But alternatively, people in the real world may be even more obedient to authority than were the experimental subjects, for example, when they get swept up in mobs or are captivated by ideological fervor. Milgram's initial research and the many replications of it give us great insight into human behavior partly because they help identify the types of people and conditions to which the initial findings (lack of resistance to authority) can be generalized. But generalizing the results of single experiments is always risky because such research often studies a small number of people who are not selected to represent any particular population.

But what if your goal is not to learn about individuals, but about the culture or subculture in a society or group? The logic of sampling does not apply if the goal is to learn about culture that is shared across individuals:

> When people all provide the same information, it is redundant to ask a question over and over. Only enough people need to be surveyed to eliminate the possibility of errors and to allow for those who might diverge from the norm. (Heise 2010:15)

If you are trying to describe a group's or society's culture, you may choose individuals for the survey based on their knowledge of the culture, not as representatives of a population of individuals (Heise 2010:16). In this situation, what is important about the individuals surveyed is what they have in common, not their diversity.

Keep these exceptions in mind, but the main point is that social scientists rarely can skirt the problem of demonstrating the generalizability of their findings. If a small sample has been studied in an experiment or a field research project, the study should be replicated in different settings or, preferably, with a representative sample of the population to which generalizations are sought (see Exhibit 5.2). The social world and the people in it are just too diverse to be considered *identical units* in most respects. Social psychological experiments and small field studies have produced good social science, but they need to be replicated in other settings, with other subjects, to claim any generalizability. Even when we believe that we have uncovered basic social processes in a laboratory experiment or field observation, we should be very concerned with seeking confirmation in other samples and in other research.

Consider a Census

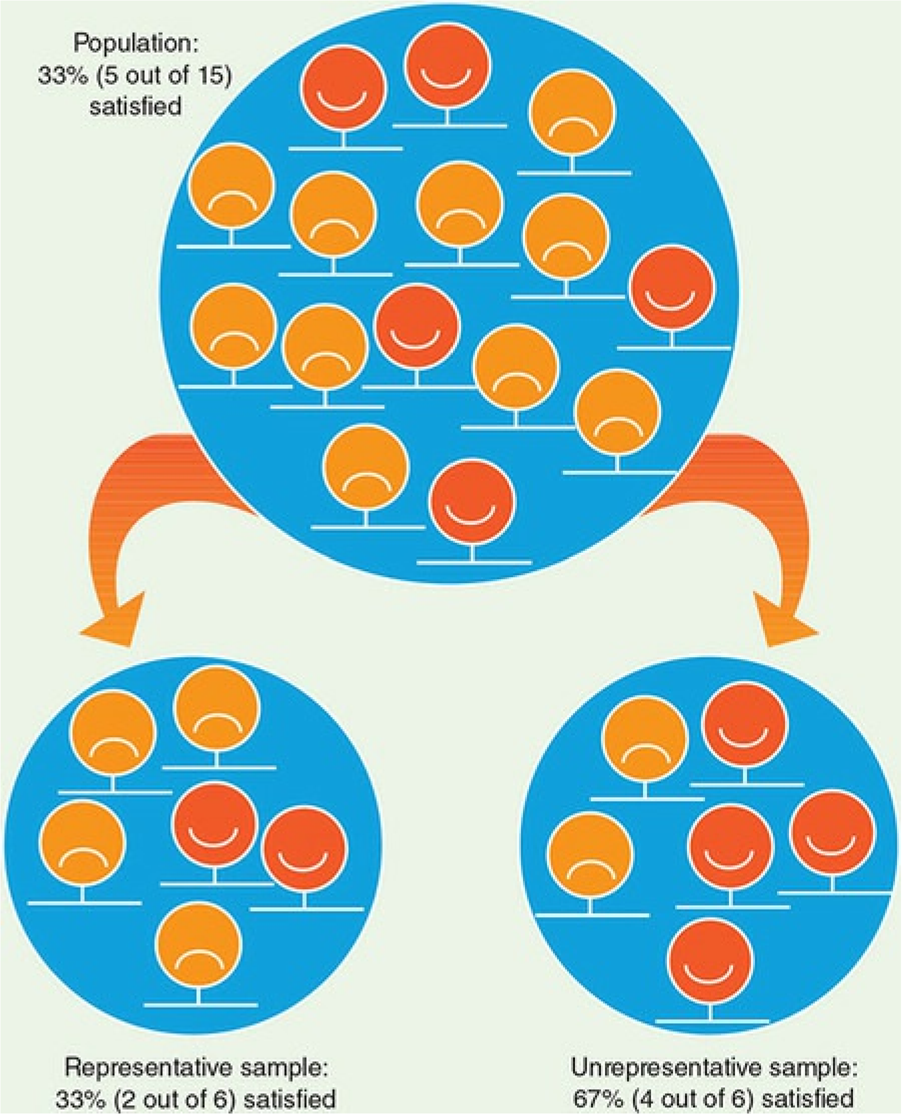
In some circumstances, it may be feasible to skirt the issue of generalizability by conducting a census—studying the entire population of interest—rather than drawing a sample. This is what the federal government tries to do every 10 years with the U.S. Census. Censuses also include studies of all the employees (or students) in small organizations, studies comparing all 50 states, and studies of the entire population of a particular type of organization in some area. However, in comparison with the U.S. Census and similar efforts in other countries, states, and cities, the population that is studied in these other censuses is relatively small.

The reason that social scientists don't often attempt to collect data from all the members of some large population is simply that doing so would be too expensive and time-consuming—and they can do almost as well with a sample. Some social scientists conduct research with data from the U.S. Census, but the government collects the data and our tax dollars pay for the effort to get one person in about 134 million households to answer 10 questions. To conduct the 2010 census, the U.S. Census Bureau spent more than $5.5 billion and hired 3.8 million people (U.S. Census Bureau 2010a, 2010b).

Representative sample: A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Census: Research in which information is obtained through responses from or information about all available members of an entire population.

Exhibit 5.2 Representative and Unrepresentative Samples



Even if the population of interest for a survey is a small town of 20,000 or students in a university of 10,000, researchers will have to sample. The costs of surveying "just" thousands of individuals exceed by far the budgets for most research projects. In fact, not even the U.S. Census Bureau can afford to have everyone answer all the questions that should be covered in the census. So it draws a sample. Every household must complete a short version of the census (it had 10 basic questions in 2010), but a sample of 3 million households is sent a long form (with about 60 questions) every year (U.S. Census Bureau 2010d). This more detailed sample survey was launched in 2005 as the American Community Survey and replaces what formerly was a long form of the census that was administered to one sixth of the population at the same time as the regular census. The 2020 Census will be distributed online in an effort to save even more money, and there are concerns that insufficient funds will be allocated to reach those who do not respond online (Wines 2017).

The fact that it is hard to get people to complete a survey is another reason why survey research can be costly. Even the U.S. Census Bureau (1999) must make multiple efforts to increase the rate of response despite the federal law requiring all citizens to complete their census questionnaire. Almost three quarters (72%) of the U.S. population returned their 2010 census questionnaire through the mail (costing 42 cents per envelope) (U.S. Census Bureau 2010a, 2010c). However, 565,000 temporary workers and as many as six follow-ups were required to contact the rest of the households that did not respond by mail, at a cost of $57 per nonrespondent (U.S. Census Bureau 2010a, 2010c). Even after all that, we know from the 2000 U.S. Census that some groups are still likely to be underrepresented (Armas 2002; Holmes 2001a), including minority groups (Kershaw 2000), impoverished cities (Zielbauer 2000), well-to-do individuals in gated communities and luxury buildings (Langford 2000), and even college students (Abel 2000). The number of persons missed in the 2000 census was estimated to be between 3.2 and 6.4 million (U.S. Census Bureau 2001).

The average survey project has far less legal and financial backing, and thus an adequate census is not likely to be possible. Consider the problems of conducting a census in Afghanistan. The first census in 23 years was conducted by the country's Central Statistics Office in 2003 and 2004, interrupted by snow that cut off many districts for 6 or 7 months. Teams of census takers carried tents, sleeping bags, and satellite phones as they trekked into remote mountainous provinces. An accompanying cartographer identified the location of each village using global positioning systems (GPS) (Gall 2003:A4). Even in Russia, which spent almost $200 million to survey its population of about 145 million, resource shortages after the collapse of the Soviet Union prevented an adequate census (Myers 2002). In Vladivostok, "Many residents, angry about a recent rise in electricity prices, refused to take part. Residents on Russian Island . . . boycotted to protest dilapidated roads" (Tavernise 2002:A13). In Iraq, dominant groups may have delayed conducting a census for fear that it would document gains in population among disadvantaged groups and thereby strengthen their claims for more resources (Myers 2010:A10).

In most survey situations, it is much better to survey only a limited number from the total population so that there are more resources for follow-up procedures that can overcome reluctance or indifference about participation. (I give more attention to the problem of nonresponse in Chapter 8.)

Sampling Methods

We can now study more systematically the features of samples that make them more or less likely to represent the population from which they are selected. The most important distinction that needs to be made about the samples is whether they are based on a probability or a nonprobability sampling method. Sampling methods that allow us to know in advance how likely it is that any element of a population will be selected for the sample are termed probability sampling methods. Sampling methods that do not let us know in advance the likelihood of selecting each element are termed nonprobability sampling methods.

Probability sampling methods rely on a random, or chance, selection procedure, which is, in principle, the same as flipping a coin to decide which of two people "wins" and which one "loses." Heads and tails are equally likely to turn up in a coin toss, so both persons have an equal chance of winning. That chance, their probability of selection, is 1 out of 2, or .5.

Flipping a coin is a fair way to select one of two people because the selection process harbors no systematic bias. You might win or lose the coin toss, but you know that the outcome was due simply to chance, not to bias. For the same reason, a roll of a six-sided die is a fair way to choose one of six possible outcomes (the odds of selection are 1 out of 6, or .17). Dealing out a hand after shuffling a deck of cards is a fair way to allocate sets of cards in a poker game (the odds of each person getting a particular outcome, such as a full house or a flush, are the same). Similarly, state lotteries use a random process to select winning numbers. Thus, the odds of winning a lottery, the probability of selection, are known, even though they are very much smaller (perhaps 1 out of 1 million) than the odds of winning a coin toss.

Probability sampling method: A sampling method that relies on a random, or chance, selection method so that the probability of selection of population elements is known.

Nonprobability sampling method: A sampling method in which the probability of selection of population elements is unknown.

Probability of selection: The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all elements of a population, the probability that any particular element will be selected is 1.0. If half the elements in the population are sampled on the basis of chance (say, by tossing a coin), the probability of selection for each element is one half, or .5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

There is a natural tendency to confuse the concept of random sampling, in which cases are selected only on the basis of chance, with a haphazard method of sampling. On first impression, “leaving things up to chance” seems to imply not exerting any control over the sampling method. But to ensure that nothing but chance influences the selection of cases, the researcher must proceed methodically, leaving nothing to chance except the selection of the cases themselves. The researcher must follow carefully controlled procedures if a purely random process is to occur. When reading about sampling methods, do not assume that a random sample was obtained just because the researcher used a random selection method at some point in the sampling process. Look for those two particular problems: selecting elements from an incomplete list of the total population and failing to obtain an adequate response rate.

If the sampling frame is incomplete, a sample selected randomly from that list will not really be a random sample of the population. You should always consider the adequacy of the sampling frame. Even for a simple population such as a university’s student body, the registrar’s list is likely to be at least a bit out-of-date at any given time. For example, some students will have dropped out, but their status will not yet be officially recorded. Although you may judge the amount of error introduced in this particular situation to be negligible, the problems are greatly compounded for a larger population. The sampling frame for a city, state, or nation is always likely to be incomplete because of constant migration into and out of the area. Even unavoidable omissions from the sampling frame can bias a sample against particular groups within the population.

An inclusive sampling frame may still yield systematic bias if many sample members cannot be contacted or refuse to participate. Nonresponse is a major hazard in survey research because nonrespondents are likely to differ systematically from those who take the time to participate. You should not assume that findings from a randomly selected sample will be generalizable to the population from which the sample was selected if the rate of nonresponse is considerable (certainly not if it is much above 30%).

Probability Sampling Methods

Probability sampling methods are those in which the probability of selection is known and is not zero (so there is some chance of selecting each element). These methods randomly select elements and therefore have no systematic bias; nothing but chance determines which elements are included in the sample. This feature of probability samples makes them much more desirable than nonprobability samples when the goal is to generalize to a larger population.

Although a random sample has no systematic bias, it will certainly have some sampling error resulting from chance. The probability of selecting a head is .5 in a single toss of a coin and in 20, 30, or however many tosses of a coin you like. But it is perfectly possible to toss a coin twice and get a head both times. The random "sample" of the two sides of the coin is selected in an unbiased fashion, but it still is unrepresentative. Imagine selecting randomly a sample of 10 people from a population comprising 50 men and 50 women. Just by chance, can't you imagine finding that these 10 people include 7 women and only 3 men? Fortunately, we can determine mathematically the likely degree of sampling error in an estimate based on a random sample (as we'll discuss later in this chapter)—assuming that the sample's randomness has not been destroyed by a high rate of nonresponse or by poor control over the selection process.

Random sampling: A method of sampling that relies on a random, or chance, selection method so that every element of the sampling frame has a known probability of being selected.

Nonrespondents: People or other entities who do not participate in a study although they are selected for the sample.

Systematic bias: Overrepresentation or underrepresentation of some population characteristics in a sample resulting from the method used to select the sample; a sample shaped by systematic sampling error is a biased sample.

In general, both the size of the sample and the homogeneity (sameness) of the population affect the degree of error as a result of chance; the proportion of the population that the sample represents does not. To elaborate,

- *The larger the sample, the more confidence we can have in the sample's representativeness.* If we randomly pick 5 people to represent the entire population of our city, our sample is unlikely to be very representative of the entire population in age, gender, race, attitudes, and so on. But if we randomly pick 100 people, the odds of having a representative sample are much better; with a random sample of 1,000, the odds become very good indeed.

- *The more homogeneous the population, the more confidence we can have in the representativeness of a sample of any particular size.* Let's say we plan to draw samples of 50 from each of two communities to estimate mean family income. One community is quite diverse, with family incomes varying from $12,000 to $85,000. In the other, more homogeneous community, family incomes are concentrated in a narrow range, from $41,000 to $64,000. The estimated mean family income based on the sample from the homogeneous community is more likely to be representative than is the estimate based on the sample from the more heterogeneous community. With less variation to represent, fewer cases are needed to represent the homogeneous community.
- *The fraction of the total population that a sample contains does not affect the sample's representativeness unless that fraction is large.* We can regard any sampling fraction less than 2% with about the same degree of confidence (Sudman 1976:184). Actually, sample representativeness is not likely to increase much until the sampling fraction is quite a bit higher. Other things being equal, a sample of 1,000 from a population of 1 million (with a sampling fraction of 0.001, or 0.1%) is much better than a sample of 100 from a population of 10,000 (although the sampling fraction for this smaller sample is 0.01, or 1%, which is 10 times higher). The size of the samples is what makes representativeness more likely, not the proportion of the whole that the sample represents.

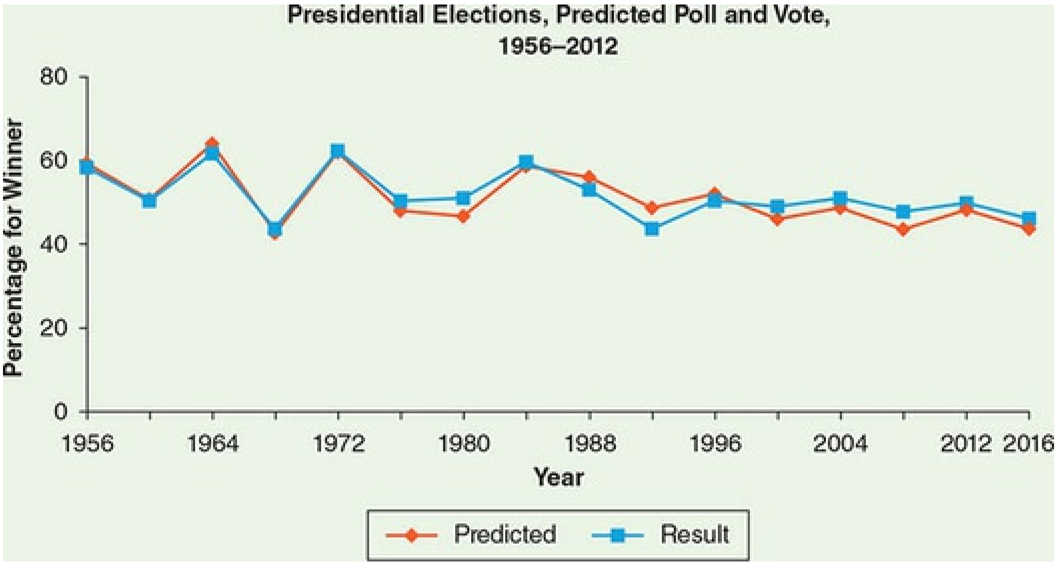
Polls to predict presidential election outcomes illustrate both the value of random sampling and the problems that it cannot overcome. Prior to the use of random sampling in polling, there were some major snafus. In 1936, a *Literary Digest* poll predicted that Alfred M. Landon would defeat President Franklin Delano Roosevelt in a landslide, but instead Roosevelt took 63% of the popular vote. The problem? The *Digest* mailed out 10 million mock ballots to people listed in telephone directories, automobile registration records, voter lists, and so on. But in 1936, during the Great Depression, only relatively wealthy people had phones and cars, and they were more likely to be Republican. Furthermore, only 2,376,523 completed ballots were returned, and a response rate of only 24% leaves much room for error. Of course, this poll was not designed as a random sample, so the appearance of systematic bias is not surprising. Gallup predicted the 1936 election results accurately with a more systematically selected sample of just 3,000 that avoided so much bias (although they did not yet use random sampling) (Bainbridge 1989:43–44).

In 1948, pollsters mistakenly predicted that Thomas E. Dewey would beat Harry S. Truman, based on the sampling method that George Gallup had used successfully since 1934. The problem was that pollsters stopped collecting data several weeks before the election, and in those weeks, many people changed their minds (Kenney 1987). The sample was systematically biased by underrepresenting shifts in voter sentiment just before the election. This experience convinced Gallup to use only random sampling methods (as well as to continue polling until the election).

In most subsequent presidential elections, pollsters have predicted accurately the outcomes of the actual votes by using random sampling and phone interviewing to learn for which candidate the likely voters intend to vote. Exhibit 5.3 shows how close these sample-based predictions have been in the past 15 contests. The exceptions were the 1980 and 1992 elections, when third-party candidates had an unpredicted effect. Otherwise, the small discrepancies between the votes predicted through random sampling and the actual votes can be attributed to random error.

The Gallup poll did not do as well in predicting the results of the 2008 presidential election. The final 2008 Gallup prediction was that Barack Obama would win with 55% to John McCain's 44% (Gallup 2011). The race turned out a bit closer, with Obama winning by 53% to McCain's 46%, with other polling organizations closer to the final mark (the Rasmussen and Pew polls were exactly on target) (Panagopoulos 2008). However, the overall rate of accuracy has been impressive.

Exhibit 5.3 Presidential Election Outcomes: Predicted and Actual



Sources: Gallup. 2011. Election Polls—Accuracy Record in Presidential Elections; Panagopoulos, Costas. 2008. "Poll Accuracy in the 2008 Presidential Election"; Jones, Jeffrey. 2012. "Gender Gap in 2012 Vote Is Largest in Gallup's History."

But what about the 2016 presidential election, with Donald Trump's surprising win (relative to expectations based on most polls)? Actually, the composite prediction from 13 national polls in the week before the election was that Hillary Clinton would win the popular vote by 3.1 points—a gap of only 1 point from her advantage of 2.1% in the national vote total calculated in November (Newport 2016). (Gallup had decided not to conduct presidential polls for the 2016 election—perhaps due to their difficulties in the 2012 election poll, although their director said it was due to their desire to survey about "issues" rather than do "horse race polling") (Clement and Craighill 2015). What polling got wrong were predictions in the so-called battleground states in which undecided voters broke for Trump in the last days and Trump supporters turned out at higher than predicted rates, while pollsters didn't adequately take account of educational differences in political preferences when they adjusted for likely turnout (college-educated voters favored Clinton by a 25-point margin) (Cohn 2017). The relatively accurate prediction from national polls of Clinton's popular margin therefore did not accurately forecast the large advantage for Trump in the electoral college due to these battleground states.

Because they do not disproportionately exclude or include particular groups within the population, random samples that are implemented successfully avoid systematic bias in the selection process. However, when some types of people are more likely to refuse to participate in surveys or are less likely to be available for interviews, or are less likely to disclose their sentiments, systematic bias can still creep into the sampling process. In addition, random error will still influence the specific results obtained from any random sample and opinions that are in flux necessarily create uncertainty.

The likely amount of random error will also vary with the specific type of random sampling method used, as I explain in the next sections. The four most common methods for drawing random samples are (1) simple random sampling, (2) systematic random sampling, (3) stratified random sampling, and (4) cluster sampling.

Simple Random Sampling

Simple random sampling requires some procedure that generates numbers or otherwise identifies cases strictly on the basis of chance. As you know, flipping a coin or rolling a die can be used to identify cases strictly on the basis of chance, but these procedures are not very efficient tools for drawing a sample. A random number table, such as the one in Appendix C, simplifies the process considerably. The researcher numbers all the elements in the sampling frame and then uses a systematic procedure for picking corresponding numbers from the random number table. (Practice Exercise 1 at the end of this chapter explains the process step by step.) Alternatively, a researcher may use a lottery procedure in which each case number is written on a small card, and then the cards are mixed up and the sample is selected from the cards.

When a large sample must be generated, these procedures are cumbersome. Fortunately, a computer program can easily generate a random sample of any size (actually, most computer programs use a process that generates what is called a "pseudorandom" sequence of numbers that is not exactly the same as numbers generated by a purely chance process, but that difference has no practical effect in social science research). The researcher must first number all the elements to be sampled (the sampling frame) and then run the computer program to generate a random selection of the numbers within the desired range. The elements represented by these numbers are the sample.

Organizations that conduct phone surveys often draw random samples using another automated procedure, called random digit dialing. A machine dials random numbers within the phone prefixes corresponding to the area in which the survey is to be conducted. Random digit dialing is particularly useful when a sampling frame is not available. The researcher simply replaces any inappropriate number (e.g., those that are no longer in service or that are for businesses) with the next randomly generated phone number.

As the fraction of the population that has only cell phones has increased (40% in 2013), it has become essential to explicitly sample cell phone numbers as well as landline phone numbers (McGeeney and Keeter 2014). Compared with those who have a landline phone, those who use cell phones only tend to be younger; are more likely to be male, single, and black or Hispanic; and are less likely to vote. As a result, failing to include cell phone numbers in a phone survey can introduce bias (Christian et al. 2010). In fact, in a 2008 presidential election survey, those who use only cell phones were less likely to be registered voters than were landline users but were considerably more favorable to Obama than landline users (Keeter 2008) (see Exhibit 5.4).

The probability of selection in a true simple random sample is equal for each element. If a sample of 500 is selected from a population of 17,000 (i.e., a sampling frame of 17,000), then the probability of selection for each element is 500 to 17,000, or .03. Every element has an equal chance of being selected, just like the odds in a toss of a coin (1 to 2) or a roll of a die (1 to 6). Thus, simple random sampling is an *equal probability of selection method,* or EPSEM.

Simple random sampling can be done either with or without replacement sampling. In replacement sampling, each element is returned to the sampling frame after it is selected so that it may be sampled again. In sampling without replacement, each element selected for the sample is then excluded from the sampling frame. In practice, it makes no difference whether sampled elements are replaced after selection as long as the population is large and the sample is to contain only a small fraction of the population. Random sampling with replacement is, in fact, rarely used.

Simple random sampling: A method of sampling in which every sample element is selected only on the basis of chance, through a random process.

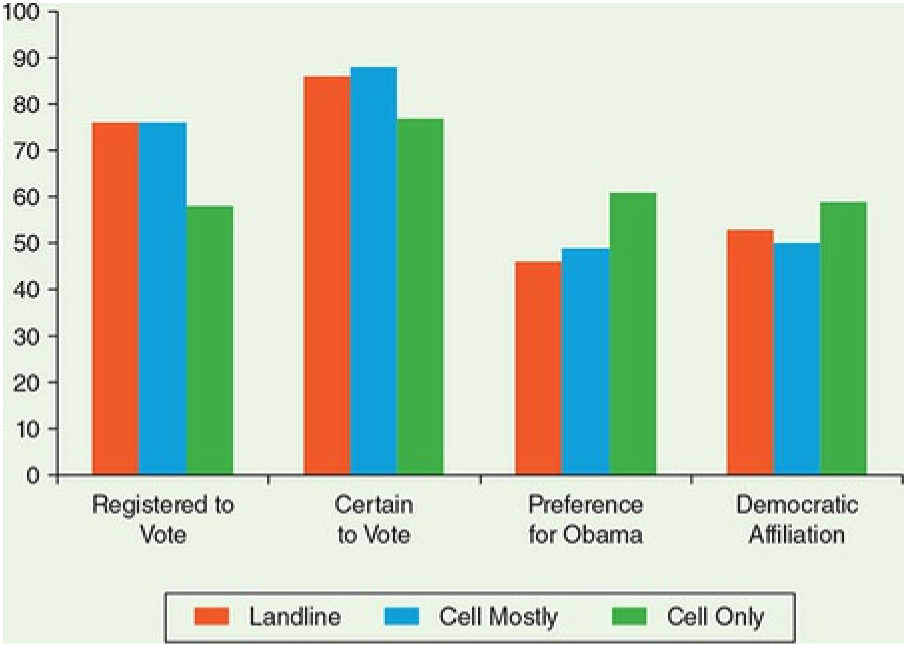
Random number table: A table containing lists of numbers that are ordered solely on the basis of chance; it is used for drawing a random sample.

Random digit dialing: The random machine-dialing of numbers within designated phone prefixes, which creates a random sample for phone surveys.

Replacement sampling: A method of sampling in which sample elements are returned to the sampling frame after being selected, so they may be sampled again. Random samples may be selected with or without replacement.

In a study involving simple random sampling, Bruce Link and his associates (1996) used random digit dialing to contact adult household members in the continental United States for an investigation of public attitudes and beliefs about homeless people. Of the potential interviewees, 63% responded. The sample obtained was not exactly comparable with the population sampled: Compared with U.S. Census figures, the sample overrepresented women, people aged 2554, married people, and those with more than a high school education; it underrepresented Latinos.

Exhibit 5.4 Cell-Only and Landline Users in a 2008 Presidential Poll



Source: Based on Keeter, Scott, Michael Dimock, and Leah Christian. “Cell Phones and the 2008 Vote: An Update.” Pew Center for the People & the Press, September 23, 2008.

How does this sample strike you? Let’s assess sample quality using the questions posed earlier in the chapter:

- *From what population were the cases selected?* There is a clearly defined population: the adult residents of the continental United States (who live in households with phones).
- *What method was used to select cases from this population?* The case selection method is a random selection procedure, and there are no systematic biases in the sampling.
- *Do the cases that were studied represent, in the aggregate, the population from which they were selected?* The findings will very likely represent the population sampled because there were no biases in the sampling and a very large number of cases were selected. However, 37% of those selected for interviews could not be contacted or chose not to

respond. This rate of nonresponse seems to create a small bias in the sample for several characteristics.

We must also consider the issue of cross-population generalizability: Do findings from this sample have implications for any larger group beyond the population from which the sample was selected? Because a representative sample of the entire U.S. adult population was drawn, this question has to do with cross-national generalizations. Link and his colleagues (1996) don't make any such generalizations. There's no telling what might occur in other countries with different histories of homelessness and different social policies.

Systematic Random Sampling

Systematic random sampling is a variant of simple random sampling. The first element is selected randomly from a list or from sequential files, and then every *n*th element is selected. This is a convenient method for drawing a random sample when the population elements are arranged sequentially. It is particularly efficient when the elements are not actually printed (i.e., there is no sampling frame) but instead are represented by folders in filing cabinets.

Systematic random sampling requires the following three steps:

1. The total number of cases in the population is divided by the number of cases required for the sample. This division yields the sampling interval, the number of cases from one sampled case to another. If 50 cases are to be selected out of 1,000, the sampling interval is 20; every 20th case is selected.
2. A number from 1 to 20 (or whatever the sampling interval is) is selected randomly. This number identifies the first case to be sampled, counting from the first case on the list or in the files.
3. After the first case is selected, every *n*th case is selected for the sample, where *n* is the sampling interval. If the sampling interval is not a whole number, the size of the sampling interval is varied systematically to yield the proper number of cases for the sample. For example, if the sampling interval is 30.5, the sampling interval alternates between 30 and 31. In almost all sampling situations, systematic random sampling yields what is essentially a simple random sample. The exception is a situation in which the sequence of elements is affected by periodicity—that is, the sequence varies in some regular, periodic pattern. For example, the houses in a new development with the same number of houses in each block (e.g., 8) may be listed by block, starting with the house in the northwest corner of each block and continuing clockwise. If the sampling interval is 8, the same as the periodic pattern, all the cases selected will be in the same position (see Exhibit 5.5). But in reality, periodicity and the sampling interval are rarely the same.

Stratified Random Sampling

Although all probability sampling methods use random sampling, some add steps to the sampling process to make sampling more efficient or easier. Stratified random sampling uses information known about the total population before sampling to make the sampling process more efficient. First, all elements in the population (i.e., in the sampling frame) are distinguished according to their value on some relevant characteristic. This characteristic might be year in school in a study of students, marital status in a study of family relations, or average property value in a study of towns. That characteristic forms the sampling strata, but of course you can use this approach only if you know the value of all elements in the population on this characteristic before you draw the sample. Next, elements are sampled randomly from within these strata. For example, race may be the basis for distinguishing individuals in some population of interest. Within each racial category, individuals are then sampled randomly. Of course, using this method requires more information before sampling than is the case with simple random sampling. It must be possible to categorize each element in one and only one stratum, and the size of each stratum in the population must be known.

Systematic random sampling: A method of sampling in which sample elements are selected from a list or from sequential files, with every *n*th element being selected after the first element is selected randomly within the first interval.

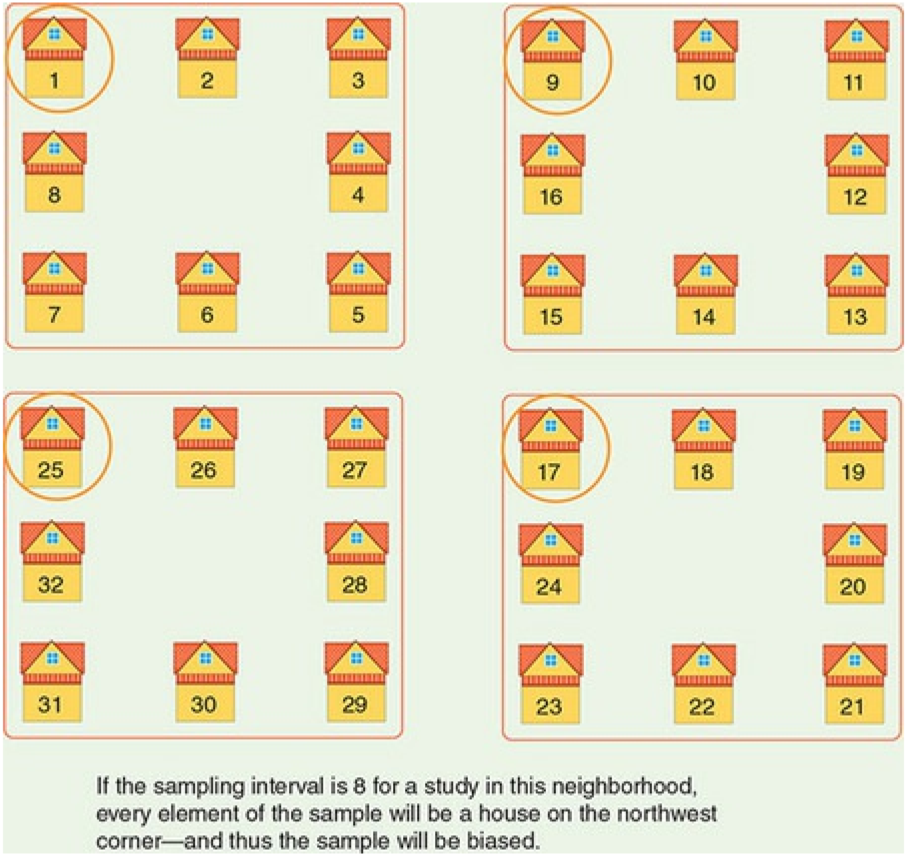
Sampling interval: The number of cases from one sampled case to another in a systematic random sample.

Periodicity: A sequence of elements (in a list to be sampled) that varies in some regular, periodic pattern.

Stratified random sampling: A sampling method in which sample elements are selected separately from population strata that are identified in advance by the researcher.

This method is more efficient than drawing a simple random sample because it ensures appropriate representation of elements across strata. Imagine that you plan to draw a sample of 500 from the population of a large company to study the experiences of different ethnic groups. You know from company records that the workforce is 15% black, 10% Hispanic, 5% Asian, and 70% white. If you drew a simple random sample, you might end up with somewhat disproportionate numbers of each group. But if you created sampling strata based on race and ethnicity, you could randomly select cases from each stratum: 75 blacks (15% of the sample), 50 Hispanics (10%), 25 Asians (5%), and 350 whites (70%). By using proportionate stratified sampling, you would eliminate any possibility of sampling error in the sample's distribution of ethnicity. Each stratum would be represented exactly in proportion to its size in the population from which the sample was drawn (see Exhibit 5.6).

Exhibit 5.5 The Effect of Periodicity on Systematic Random Sampling


1
2
3
8
4
7
6
5
9
10
11
16
12
15
14
13
25
26
27
32
28
31
30
29
17
18
19
24
20
23
22
21
If the sampling interval is 8 for a study in this neighborhood, every element of the sample will be a house on the northwest corner—and thus the sample will be biased.

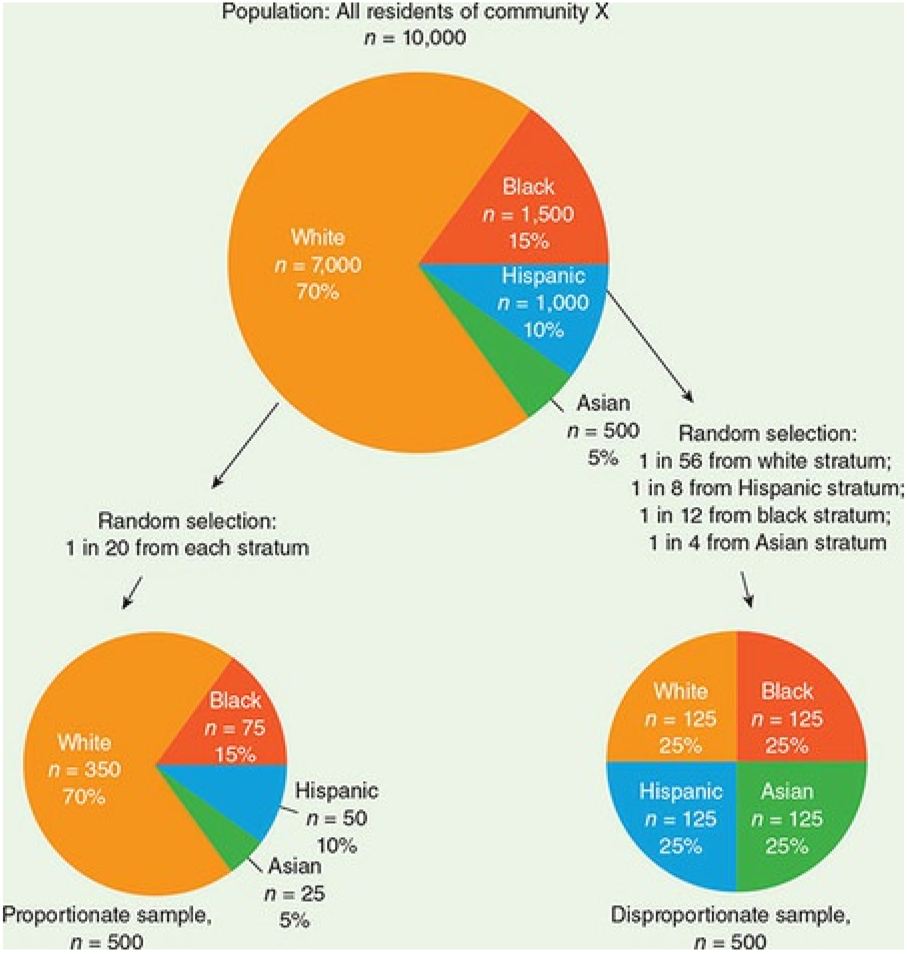
This is the strategy Brenda Booth et al. (2002) used in a study of homeless adults in two Los Angeles County sites with large homeless populations. Specifically, Booth et al. (2002:432) selected subjects at random from homeless shelters, from meal facilities, and from literally homeless populations on the streets. Respondents were sampled proportionately to their numbers in the downtown and Westside areas, as determined by a one-night enumeration. They were also sampled proportionately to their distribution across three nested sampling strata: the population using shelter beds, the population using meal facilities, and the unsheltered population using neither.

Proportionate stratified sampling: A sampling method in which elements are selected from strata in exact proportion to their representation in the population.

In disproportionate stratified sampling, the proportion of each stratum that is included in the sample is intentionally varied from what it is in the population. In the case of the company sample stratified by ethnicity, you might select equal numbers of cases from each racial or ethnic group: 125 blacks (25% of the sample), 125 Hispanics (25%), 125 Asians (25%), and 125 whites (25%). In this type of sample, the probability of selection of every case is known but unequal between strata. You know what the proportions are in the population, and so you can easily adjust your combined sample statistics to reflect these true proportions. For instance, if you want to combine the ethnic groups and estimate the average income of the total population, you would have to *weight* each case in the sample. The weight is a number you multiply by the value of each case based on the stratum it is in. For example, you would multiply the incomes of all blacks in the sample by 0.6 (75/125), the incomes of all Hispanics by 0.4 (50/125), and so on. Weighting in this way reduces the influence of the oversampled strata and increases the influence of the undersampled strata to what they would have been if pure probability sampling had been used.

Disproportionate stratified sampling: A sampling method in which elements are selected from strata in different proportions from those that appear in the population.

Exhibit 5.6 Stratified Random Sampling



Booth et al. (2002:432) included one element of *dis*proportionate random sampling in their otherwise proportionate random sampling strategy for homeless persons in Los Angeles: The researchers oversampled women so that women composed 26% of the sample compared with their actual percentage of 16% in the homeless population. Why would anyone select a sample that is so unrepresentative in the first place? The most common reason is to ensure that cases from smaller strata are included in the sample in sufficient numbers to allow separate statistical estimates and to facilitate comparisons between strata. Remember that one of the determinants of sample quality is sample size. The same is true for subgroups within samples. If a key concern in a research project is to describe and compare the incomes of people from different racial and ethnic groups, then it is important that the researchers base the mean income of each group on enough cases to be a valid representation. If few members of a particular minority group are in the population, they need to be oversampled. Such disproportionate sampling may also result in a more efficient sampling design if the costs of data collection differ markedly between the strata or if the variability (heterogeneity) of the strata differs.

Weighting is also sometimes used to reduce the lack of representativeness of a sample that occurs because of nonresponse. On finding that the obtained sample does not represent the population for some known characteristics such as, perhaps, gender or education, the researcher weights the cases in the sample so that the sample has the same proportions of men and women, or high school graduates and college graduates, as the complete population (see Exhibit 5.7). Keep in mind, though, that this procedure does not solve the problems caused by an unrepresentative sample because you still don't know what the sample composition should have been relative to the other variables in your study; all you have done is to reduce the sample's unrepresentativeness relative to the variables used in weighting. This may, in turn, make it more likely that the sample is representative of the population relative to other characteristics, but you don't really know.

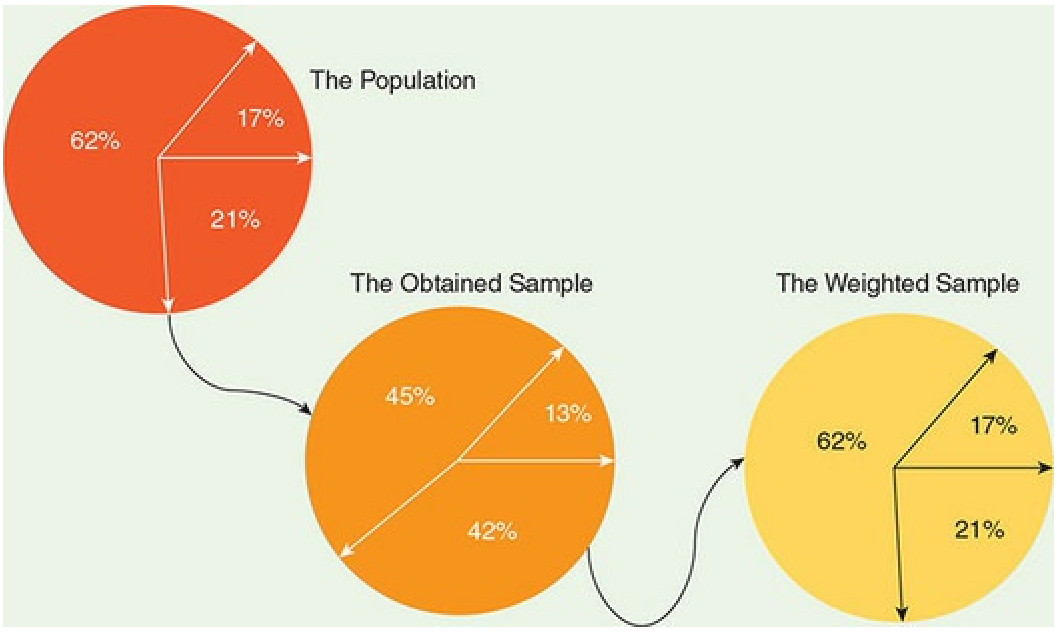
Multistage Cluster Sampling

Multistage cluster sampling is useful when a sampling frame of elements is not available, as often is the case for large populations spread out across a wide geographic area or among many different organizations. A cluster is a naturally occurring, mixed aggregate of elements of the population, with each element appearing in one, and only one, cluster. Schools could serve as clusters for sampling students, blocks could serve as clusters for sampling city residents, counties could serve as clusters for sampling the general population, and businesses could serve as clusters for sampling employees.

Multistage cluster sampling: A sampling method in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Cluster: A naturally occurring, mixed aggregate of elements of the population.

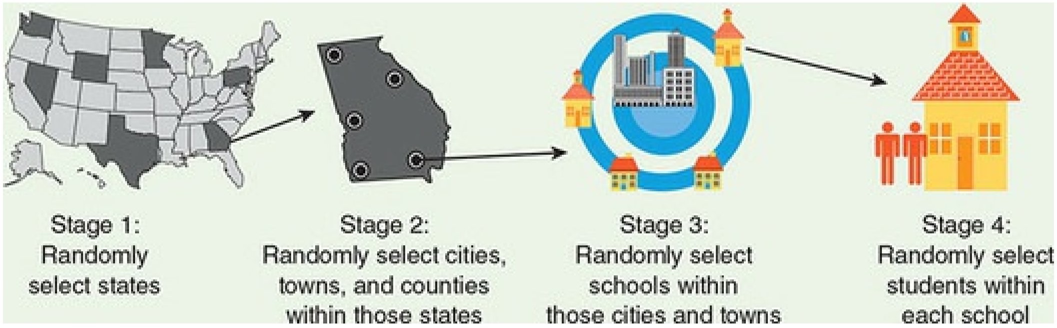
Exhibit 5.7 Weighting an Obtained Sample to Match a Population Proportion



Drawing a multistage cluster sample is, at least, a two-stage procedure. First, the researcher draws a random sample of clusters. A list of clusters should be much easier to obtain than a list of all the individuals in each cluster in the population. Next, the researcher draws a random sample of elements within each selected cluster. Because only a fraction of the total clusters are involved, obtaining the sampling frame at this stage should be much easier.

In a cluster sample of city residents, for example, blocks could be the first-stage clusters. A research assistant could walk around each selected block and record the addresses of all occupied dwelling units. Or, in a cluster sample of students, a researcher could contact the schools selected in the first stage and make arrangements with the registrar to obtain lists of students at each school. Cluster samples often involve more than two stages (see Exhibit 5.8), with clusters within clusters, as when a national sample of individuals might involve first sampling states, then geographic units within those states, then dwellings within those units, and finally, individuals within the dwellings. In multistage cluster sampling, the clusters at the first stage of sampling are termed the *primary sampling units* (Levy and Lemeshow 1999:228).

Exhibit 5.8 Multistage Cluster Sampling


Stage 1:
Randomly
select states
Stage 2:
Randomly select cities,
towns, and counties
within those states
Stage 3:
Randomly select
schools within
those cities and towns
Stage 4:
Randomly select
students within
each school

How many clusters should be selected, and how many individuals within each cluster should be selected? As a general rule, the sample will be more similar to the entire population if the researcher selects as many clusters as possible—even though this will mean the selection of fewer individuals within each cluster. Unfortunately, this strategy also maximizes the cost of the sample for studies using in-person interviews. The more clusters a researcher selects, the more time and money will have to be spent traveling to the different clusters to reach the individuals for interviews.

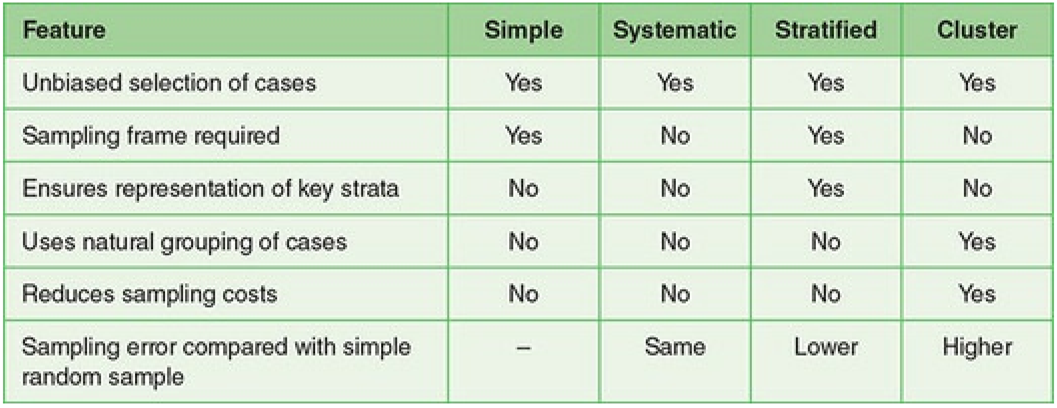
The calculation of how many clusters to sample and how many individuals are within the clusters is also affected by the degree of similarity of individuals within clusters: The more similar the individuals are within the clusters, the fewer the number of individuals needed to represent each cluster. So if you set out to draw a cluster sample, be sure to consider how similar individuals are within the clusters as well as how many clusters you can afford to include in your sample.

Multistage cluster sampling is a very popular method among survey researchers, but it has one general drawback: Sampling error is greater in a cluster sample than in a simple random sample because there are two steps involving random selection rather than just one. This sampling error increases as the number of clusters decreases, and it decreases as the homogeneity of cases per cluster increases. In sum, it's better to include as many clusters as possible in a sample, and it's more likely that a cluster sample will be representative of the population if cases are relatively similar within clusters.

Probability Sampling Methods Compared

Can you now see why researchers often prefer to draw a stratified random sample or a cluster sample rather than a simple random sample? Exhibit 5.9 should help you remember the key features of these different types of sampling and to determine when each is most appropriate.

Exhibit 5.9 Features of Probability Sampling Methods



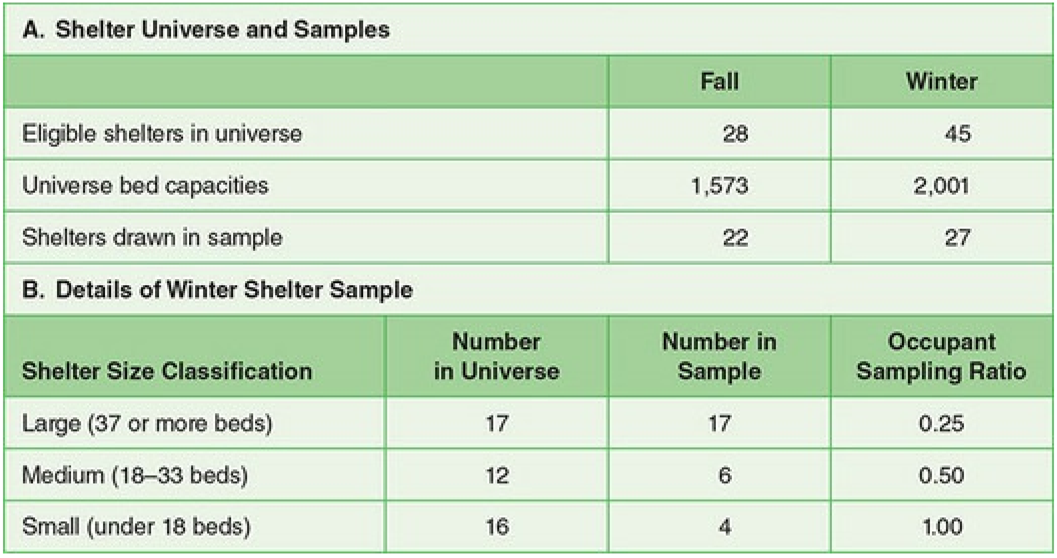
Feature	Simple	Systematic	Stratified	Cluster
Unbiased selection of cases	Yes	Yes	Yes	Yes
Sampling frame required	Yes	No	Yes	No
Ensures representation of key strata	No	No	Yes	No
Uses natural grouping of cases	No	No	No	Yes
Reduces sampling costs	No	No	No	Yes
Sampling error compared with simple random sample	–	Same	Lower	Higher

Many professionally designed surveys use combinations of clusters and stratified probability sampling methods. For example, Peter Rossi (1989) drew a disproportionate stratified cluster sample of shelter users for a homelessness study in Chicago (see Exhibit 5.10). The shelter sample was stratified by size, with smaller shelters having a smaller likelihood of selection than larger shelters. In fact, the larger shelters were all selected; they had a probability of selection of 1.0. Within the selected shelters, shelter users were then sampled using a systematic random selection procedure (except in the small shelters, in which all persons were interviewed). Homeless persons living on the streets were also sampled randomly. In the first stage, city blocks were classified into strata based on the likely concentration of homeless persons (estimated by several knowledgeable groups). Blocks were then picked randomly within these strata and, on the survey night between 1 a.m. and 6 a.m., teams of interviewers screened each person found outside on that block for his or her homeless status. Persons identified as homeless were then interviewed (and given $5 for their time). The rate of response for two different samples (fall and winter) in the shelters and on the streets was between 73% and 83%.

How would we evaluate the Chicago homeless sample, using the sample evaluation questions?

- *From what population were the cases selected?* The population was clearly defined for each cluster.
- *What method was used to select cases from this population?* The random selection method was carefully described.
- *Do the cases that were studied represent, in the aggregate, the population from which they were selected?* The unbiased selection procedures make us reasonably confident in the representativeness of the sample, although we know little about the nonrespondents and therefore may justifiably worry that some types of homeless persons were missed.

Exhibit 5.10 Chicago Shelter Universe and Shelter Samples, Fall and Winter Surveys



A. Shelter Universe and Samples

	Fall	Winter
Eligible shelters in universe	28	45
Universe bed capacities	1,573	2,001
Shelters drawn in sample	22	27

B. Details of Winter Shelter Sample

Shelter Size Classification	Number in Universe	Number in Sample	Occupant Sampling Ratio
Large (37 or more beds)	17	17	0.25
Medium (18–33 beds)	12	6	0.50
Small (under 18 beds)	16	4	1.00

Source: Rossi, Peter H. 1989. Down and Out in America: The Origins of

Homelessness. Reprinted with permission from the University of Chicago Press.

Cross-population generalization is reasonable with this sample because it seems likely that the findings reflect general processes involving homeless persons. Rossi (1989) clearly thought so because his book’s title refers to homelessness in America, not just in Chicago.

Nonprobability Sampling Methods

Nonprobability sampling methods are often used in qualitative research; they also are used in quantitative studies when researchers are unable to use probability selection methods. In qualitative research, a focus on one setting or a very small sample allows a more intensive portrait of activities and actors, and may lead to confidence in having represented fairly the entire setting and participants in it. Even when this argument seems persuasive, though, we must still be concerned with the external validity of the findings—the extent to which they can be generalized to other places. In the many studies in which qualitative researchers have interviewed only a small portion of the participants in a setting under investigation or observed just a limited portion of the events of interest, we have to realize that the use of nonprobability sampling methods limits the ability to generalize to the whole population of participants or events in the setting and lowers the confidence that others can place in these generalizations.

There are four common nonprobability sampling methods: (1) availability sampling, (2) quota sampling, (3) purposive sampling, and (4) snowball sampling. Because these methods do not use a random selection procedure, we cannot expect a sample selected with any of these methods to yield a representative sample. They should not be used in quantitative studies if a probability-based method is feasible. Nonetheless, these methods are useful when random sampling is not possible, when a research question calls for an intensive investigation of a small population, or when a researcher is performing a preliminary, exploratory study.

Availability (Convenience) Sampling

Elements are selected for availability sampling—also called "convenience sampling"—because they're available or easy to find. Thus, this sampling method is also known as haphazard or accidental sampling. There are many ways to select elements for an availability sample: standing on street corners and talking to whoever walks by, asking questions of employees who have time to talk when they pick up their paychecks at a personnel office, or approaching particular individuals at opportune times while observing activities in a social setting. You may find yourself interviewing available students at campus hangouts as part of a course assignment. A participant observation study of a group may require no more sophisticated approach. When Philippe Bourgois, Mark Lettiere, and James Quesada (1997) studied homeless heroin addicts in San Francisco, they immersed themselves in a community of addicts living in a public park. These addicts became the availability sample.

Availability sampling: A nonprobability sampling method in which elements are selected on the basis of convenience.

An availability sample is appropriate in social research when a field researcher is exploring a new setting and trying to get some sense of the prevailing attitudes or when a survey researcher conducts a preliminary test of a new set of questions, but it can also be required when the population simply cannot be listed or otherwise identified in advance—at least with available resources. It was because of the lack of an adequate sampling frame (list of the population) and sufficient resources that Kristin Ferguson, Kimberly Bender, Sanna Thompson, Elaine M. Maccio, and David Pollio (this chapter's "Research That Matters" authors) used an availability sampling strategy to select 238 homeless young people to interview in five cities. The researchers standardized their sampling methods between the five cities: They recruited at agencies serving the homeless during the same time periods, and they maintained the same eligibility requirements between cities. Before interviews began, a short eligibility screening form was used to check age and homeless status. Ferguson and her collaborators then recruited an even smaller subset of respondents in one city for an exploratory discussion of challenges in finding employment.

Now, answer the sample evaluation questions in relation to the homeless young adult interviews by Ferguson and her collaborators. If your answers are something like "The population was intended to be all homeless young adults in these five cities," "The method for selecting cases was based on availability," and "We cannot determine the extent to which the youth studied represent the population," you're right! There certainly is not much likelihood in such an availability sample that the interviewees represent the distribution of experiences among all homeless youth adults in these five cities. Of course, there's also no basis for concluding that these homeless youth are similar to those in other cities or in nonurban areas. However, if you have read much about homeless youth in other locations and have come to believe that there are many more similarities than differences between them, you might believe that influences on their employment status are likely to be similar across settings.

In a similar vein, perhaps person-in-the-street comments to news reporters suggest something about what homeless persons think, or maybe they don't; we can't really be sure. But let's give reporters their due: If they just want to have a few quotes to make their story more appealing, nothing is wrong with their sampling method. However, their approach gives us no basis for thinking that we have an overview of community sentiment. The people who happen to be available in any situation are unlikely to be just like those who are unavailable. We can't be at all certain that what we learn can be generalized with any confidence to a larger population of concern.

Availability sampling often masquerades as a more rigorous form of research. Popular magazines periodically survey their readers by printing a questionnaire for readers to fill out and mail in. A follow-up article then appears in the magazine under a title such as "What

You Think About Intimacy in Marriage." If the magazine's circulation is large, a large sample can be achieved in this way. The problem is that usually only a tiny fraction of readers return the questionnaire, and these respondents are probably unlike other readers who did not have the interest in participating or the time to do so. So the survey is based on an availability sample. Even though the follow-up article may be interesting, we have no basis for thinking that the results describe the readership as a whole—much less the population at large.

Do you see now why availability sampling differs so much from random sampling methods, which require that "nothing but chance" affects the actual selection of cases? What makes availability sampling "haphazard" is precisely that a great many things other than chance can affect the selection of cases, ranging from the prejudices of the research staff to the work schedules of potential respondents. To truly leave the selection of cases up to chance, we have to design the selection process very carefully so that other factors are not influential. There's nothing haphazard about selecting cases randomly.

Careers and Research

Tyler Evans, Account Supervisor

When Tyler Evans began his career in advertising and public relations, he did not have a particular fondness for research. He says, "I thought my 'gut' was going to help separate my ideas from the others." However, he found out quickly that relying on his gut would be no better than shooting at a target in the dark, blindfolded.

Like other marketers who have advanced in their careers, Evans learned that he had to understand the "why" before he could get to the "how." Hubris—the assumption that you "just knew" what the answer was, without research—is the biggest impediment to that understanding. Talented marketers figure out quickly that to solve the complex consumer puzzle, to learn what is behind the consumer's motivation, it is imperative to understand the consumer's need states. Few realize that the process of understanding must begin with identifying fundamental consumer truths.

The person most accountable for this is the agency lead on the account, who must understand exactly who the target is, why they are the best focus for the product or service, and what strategy will best engage them. There is a saying that hangs over many computer monitors to keep that strategic focus at the forefront each day, and it reads simply: "Data beats opinions." Research must lead the strategy and, eventually, the tactics of successful marketers. This research goes beyond simple demographic information to create a knowledge base steeped in understanding consumer need states and what can be offered during their moments of need. If there is not that deep understanding at the intersection of need states and product offerings, then consumers will not make any long-term commitment to the product; many dollars will be wasted chasing unattainable targets. To be just a bit hyperbolic, the ultimate goal in working at an advertising or PR firm is to do bold work that sends shockwaves through the heart of the target audience.

Today, Evans leads an account for the advertising agency for which he works. He advises students interested in a marketing career that research is the one thing that marketers use on a daily basis to ensure a focus on what will be measurable, significant, and impactful: "Marketers should receive training in research so that they can understand consumer insights and become an impactful force in the marketing community."

Quota Sampling

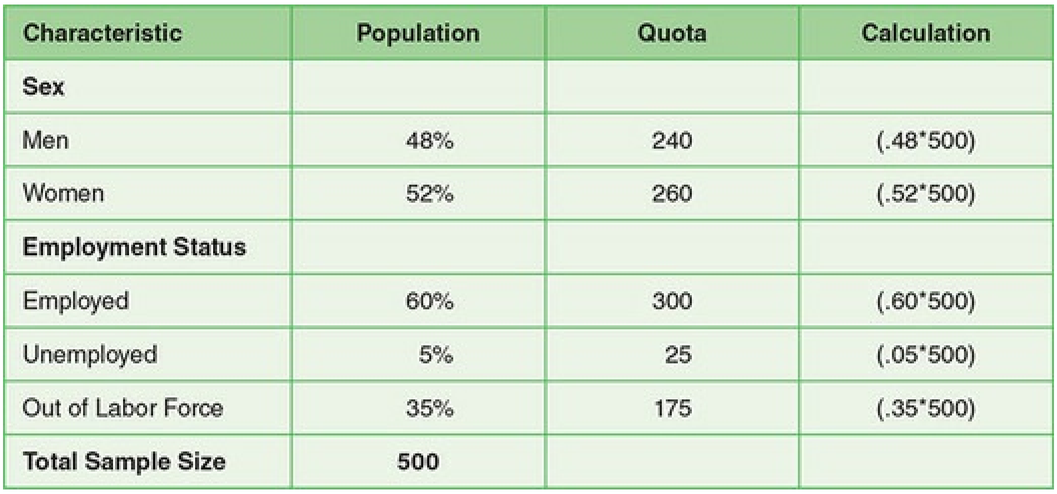
Quota sampling is intended to overcome the most obvious flaw of availability sampling—that the sample will just consist of whoever or whatever is available, without any concern for its similarity to the population of interest. The distinguishing feature of a quota sample is that quotas are set to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Quota sampling: A nonprobability sampling method in which elements are selected to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Suppose that you want to sample adult residents of a town in a study of support for a tax increase to improve the town's schools. You know from the town's annual report what the proportions of town residents are in gender, race, age, and number of children. You think that each of these characteristics might influence support for new school taxes, so you want to be sure that the sample includes men, women, whites, blacks, Hispanics, Asians, older people, younger people, big families, small families, and childless families in proportion to their numbers in the town population.

This is where quotas come in. Let's say that 48% of the town's adult residents are men and 52% are women, and that 60% are employed, 5% are unemployed, and 35% are out of the labor force. These percentages and the percentages corresponding to the other characteristics become the quotas for the sample. If you plan to include 500 residents in your sample, 240 must be men (48% of 500), 260 must be women, 300 must be employed, and so on (see Exhibit 5.11).

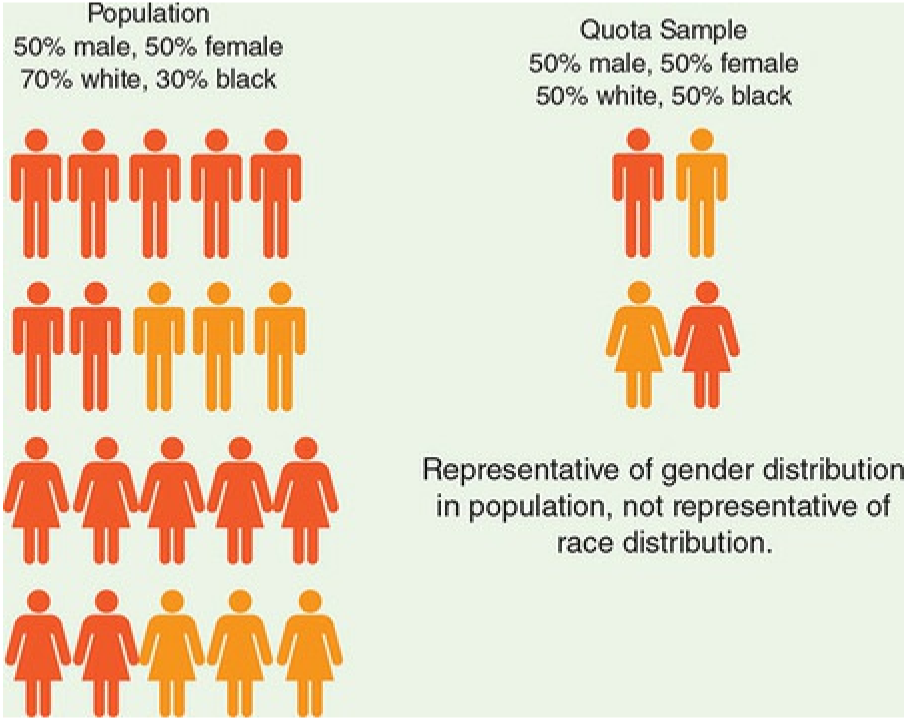
Exhibit 5.11 Adjusting a Quota Sample for Population



Characteristic	Population	Quota	Calculation
Sex			
Men	48%	240	(.48*500)
Women	52%	260	(.52*500)
Employment Status			
Employed	60%	300	(.60*500)
Unemployed	5%	25	(.05*500)
Out of Labor Force	35%	175	(.35*500)
Total Sample Size	**500**		

You may even set more refined quotas, such as certain numbers of employed women, employed men, unemployed men, and so on. With the quota list in hand, you (or your research staff) can now go out into the community looking for the right number of people in each quota category. You may go door to door or bar to bar, or just stand on a street corner until you have surveyed 240 men, 260 women, and so on.

Exhibit 5.12 Quota Sampling

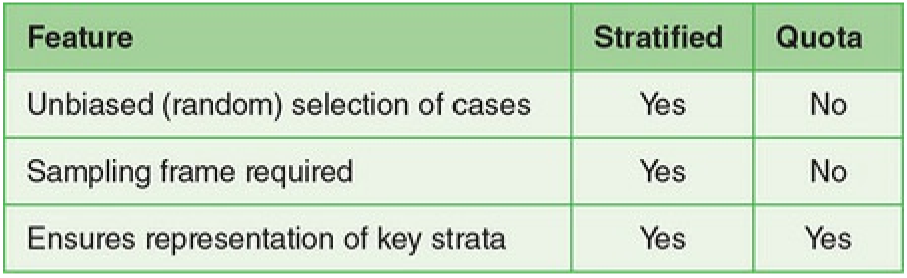

Population
50% male, 50% female
70% white, 30% black
Quota Sample
50% male, 50% female
50% white, 50% black
Representative of gender distribution in population, not representative of race distribution.

The problem is that even when we know that a quota sample is representative of the particular characteristics for which quotas have been set, we have no way of knowing if the sample is representative for any other characteristics. In Exhibit 5.12, for example, quotas have been set for gender only. Under these circumstances, it's no surprise that the sample is representative of the population only for gender, not race. Interviewers are only human; they may avoid potential respondents with menacing dogs in the front yard, or they could seek out respondents who are physically attractive or who look like they'd be easy to interview. Realistically, researchers can set quotas for only a small fraction of the characteristics relevant to a study, so a quota sample is really not much better than an availability sample (although following careful, consistent procedures for selecting cases within the quota limits always helps).

This last point leads me to another limitation of quota sampling: You must know the characteristics of the entire population to set the right quotas. In most cases, researchers know what the population looks like relative to no more than a few of the characteristics relevant to their concerns—and in some cases, they have no such information on the entire population.

If you're now feeling skeptical of quota sampling, you've gotten the drift of my remarks. Nonetheless, in some situations, establishing quotas can add rigor to sampling procedures. It's almost always better to maximize possibilities for comparison in research, and quota sampling techniques can help qualitative researchers do this. For instance, Doug Timmer, Stanley Eitzen, and Kathryn Talley (1993:7) interviewed homeless persons in several cities and other locations for their book on the sources of homelessness. Persons who were available were interviewed, but the researchers paid some attention to generating a diverse sample. They interviewed 20 homeless men who lived on the streets without shelter and 20 mothers who were found in family shelters. About half of those whom the researchers selected in the street sample were black, and about half were white. Although the researchers did not use quotas to try to match the distribution of characteristics among the total homeless population, their informal quotas helped ensure some diversity in key characteristics. Due to concerns with the value of clinical research for underrepresented groups, the National Institutes of Health now requires the inclusion of women and ethnic minorities in clinical research when appropriate and feasible (Israel 2014:44). This is another use of a quota to ensure a diverse sample.

Exhibit 5.13 Comparison of Stratified and Quota Sampling Methods



Feature	Stratified	Quota
Unbiased (random) selection of cases	Yes	No
Sampling frame required	Yes	No
Ensures representation of key strata	Yes	Yes

Does quota sampling remind you of stratified sampling? It's easy to understand why because they both select sample members partly on the basis of one or more key characteristics. Exhibit 5.13 summarizes the differences between quota sampling and stratified random sampling. The key difference, of course, is the lack of random selection in quota sampling.

Purposive Sampling

In purposive sampling, each sample element is selected for a purpose, usually because of the unique position of the sample elements. Purposive sampling may involve studying the entire population of some limited group (directors of shelters for homeless adults) or a subset of a population (mid-level managers with a reputation for efficiency). Or a purposive sample may be a *key informant survey,* which targets individuals who are particularly knowledgeable about the issues under investigation.

Purposive sampling: A nonprobability sampling method in which elements are selected for a purpose, usually because of their unique position.

Herbert Rubin and Irene Rubin (1995) suggest three guidelines for selecting informants when designing any purposive sampling strategy. Informants should be

- Knowledgeable about the cultural arena or situation or experience being studied
- Willing to talk
- Represent[ative of] the range of points of view (p. 66)

In addition, Rubin and Rubin (1995) suggest continuing to select interviewees until you can pass two tests:

> *Completeness:* What you hear provides an overall sense of the meaning of a concept, theme, or process. (p. 72)
>
> *Saturation:* You gain confidence that you are learning little that is new from subsequent interview[s]. (p. 73)

Adhering to these guidelines will help ensure that a purposive sample adequately represents the setting or issues studied.

Of course, purposive sampling does not produce a sample that represents some larger population, but it can be exactly what is needed in a case study of an organization, community, or some other clearly defined and relatively limited group. In an intensive organizational case study, a purposive sample of organizational leaders might be complemented with a probability sample of organizational members. Before designing her probability samples of hospital patients and homeless persons, Dee Roth (1990:146–147) interviewed a purposive sample of 164 key informants from organizations that had contact with homeless people in each of the counties she studied.

Snowball Sampling

Snowball sampling is useful for hard-to-reach or hard-to-identify populations for which there is no sampling frame, but the members of which are somewhat interconnected (at least some members of the population know each other). It can be used to sample members of groups such as drug dealers, prostitutes, practicing criminals, participants in Alcoholics Anonymous groups, gang leaders, informal organizational leaders, and homeless persons. It may also be used for charting the relationships between members of some group (a sociometric study), for exploring the population of interest before developing a formal sampling plan, and for developing what becomes a census of informal leaders of small organizations or communities. However, researchers using snowball sampling normally cannot be confident that their sample represents the total population of interest, so generalizations must be tentative.

Snowball sampling: A nonprobability sampling method in which sample elements are selected as they are identified by successive informants or interviewees.

Rob Rosenthal (1994) used snowball sampling to study homeless persons living in Santa Barbara, California:

> I began this process by attending a meeting of homeless people I had heard about through my housing advocate contacts. . . . One homeless woman . . . invited me to . . . where she promised to introduce me around. Thus a process of snowballing began. I gained entree to a group through people I knew, came to know others, and through them gained entree to new circles. (pp. 178, 180)

One problem with this technique is that the initial contacts may shape the entire sample and foreclose access to some members of the population of interest:

> Sat around with [my contact] at the Tree. Other people come by, are friendly, but some regulars, especially the tougher men, don't sit with her. Am I making a mistake by tying myself too closely to her? She lectures them a lot. (Rosenthal 1994:181)

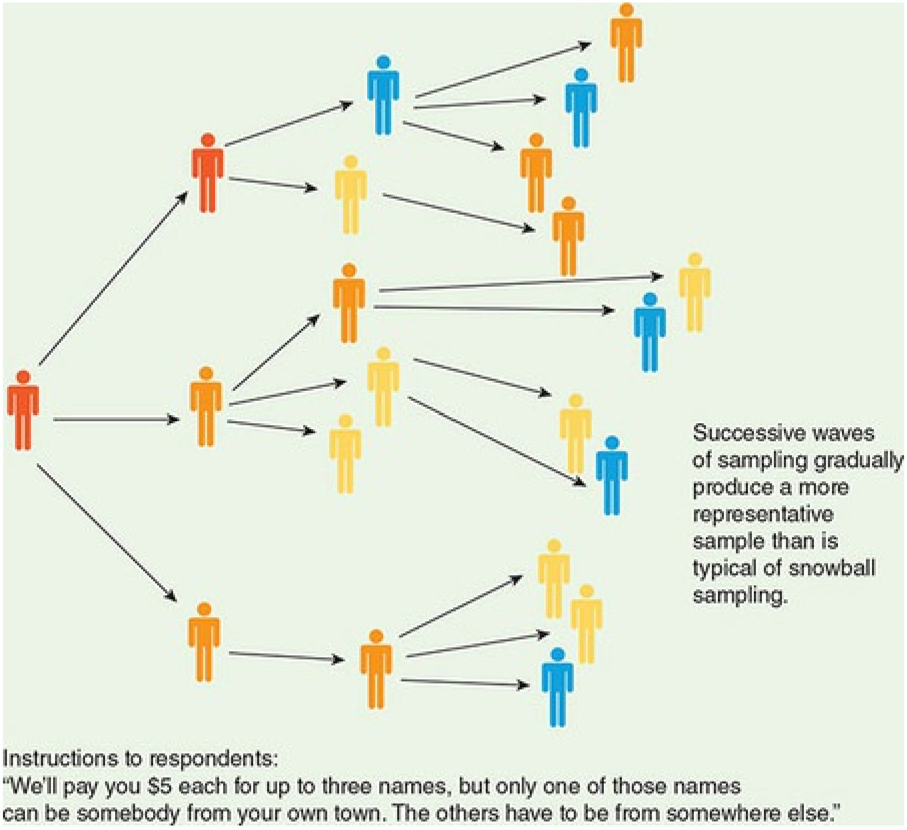
More systematic versions of snowball sampling can reduce the potential for bias. For example, *respondent-driven sampling* gives financial incentives to respondents to recruit diverse peers (Heckathorn 1997). Limitations on the number of incentives that any one respondent can receive increase the sample's diversity. Targeted incentives can steer the sample to include specific subgroups. When the sampling is repeated through several waves, with new respondents bringing in more peers, the composition of the sample converges on a more representative mix of characteristics than would occur with uncontrolled snowball sampling. Exhibit 5.14 shows how the sample spreads out through successive recruitment waves to an increasingly diverse pool (Heckathorn 1997:178). Exhibit 5.15 shows that even if the starting point were all white persons, respondent-driven sampling would result in an appropriate ethnic mix from an ethnically diverse population (Heckathorn 2002:17).

Lessons About Sample Quality

Some lessons are implicit in my evaluations of the samples in this chapter:

- We can’t evaluate the quality of a sample if we don’t know what population it is supposed to represent. If the population is unspecified because the researchers were never clear about the population they were trying to sample, then we can safely conclude that the sample itself is no good.

Exhibit 5.14 Respondent-Driven Sampling

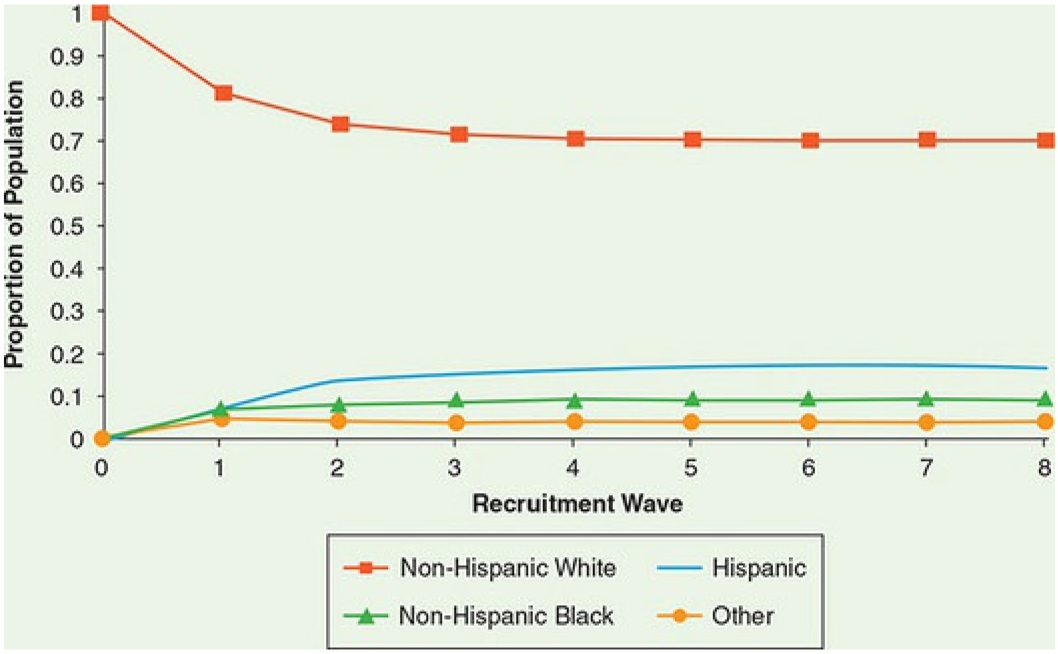


Successive waves of sampling gradually produce a more representative sample than is typical of snowball sampling.

Instructions to respondents:
“We’ll pay you $5 each for up to three names, but only one of those names can be somebody from your own town. The others have to be from somewhere else.”

Source: Based on Heckathorn (1997):178.

Exhibit 5.15 Convergence of Respondent-Driven Sample to True Ethnic Proportions in Population, After Starting With Only Whites



Source: Copyright © 2002, The Society for the Study of Social Problems Inc. Reprinted with permission from Oxford University Press.

- We can’t evaluate the quality of a sample if we don’t know how cases in the sample were selected from the population. If the method was specified, we then need to know whether cases were selected in a systematic fashion and on the basis of chance. In any case, we know that a haphazard method of sampling (as in person-on-the-street interviews) undermines generalizability.
- Sample quality is determined by the sample actually obtained, not just by the sampling method itself. If many of the people selected for our sample are nonrespondents or people (or other entities) who do not participate in the study although they have been selected for the sample, the quality of our sample is undermined—even if we chose the sample in the best possible way.
- We need to be aware that even researchers who obtain very good samples may talk about the implications of their findings for some group that is larger than, or just different from, the population they actually sampled. For example, findings from a representative sample of students in one university often are discussed as if they tell us about university students in general. And maybe they do; we just don’t know for sure.
- A sample that allows comparisons involving theoretically important variables is better than one that does not allow such comparisons. Even when we study people or social processes in depth, it is best to select individuals or settings with an eye to how useful they will be for examining relationships. Limiting an investigation to just one setting or just one type of person will inevitably leave us wondering what it is that makes a difference.

Generalizability in Qualitative Research

Qualitative research often focuses on populations that are hard to locate or very limited in size. In consequence, nonprobability sampling methods such as availability sampling and snowball sampling are often used. However, this does not mean that generalizability should be ignored in qualitative research, or that a sample should be studied simply because it is convenient (Gobo 2008:206). Janet Ward Schofield (2002) suggests two different ways of increasing the generalizability of the samples obtained in such situations:

> *Studying the Typical.* Choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience. (p. 181)
>
> *Performing Multisite Studies.* A finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites. . . . Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more . . . likely to be useful in understanding various other sites than one emerging from the study of several very similar sites. (p. 184)

Ferguson et al's. (2012) study of homeless youth is a good example of this "multisite" approach to increasing generalizability in qualitative research.

Giampietro Gobo (2008:204–205) highlights another approach to improving generalizability in qualitative research. A case may be selected for in-depth study because it is atypical, or deviant. Investigating social processes in a situation that differs from the norm will improve understanding of how social processes work in typical situations: "the exception that proves the rule."

Rather than attempting to improve generalizability, some qualitative researchers instead question its value as a goal. The argument is that understanding the particulars of a situation in depth is an important object of inquiry in itself. In the words of sociologist Norman Denzin,

> The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. . . . Every instance of social interaction . . . represents a slice from the life world that is the proper subject matter for interpretive inquiry. (Denzin cited in Schofield 2002:173)

But see if you can read the results of an intensive study of even a single setting or group without considering the extent to which the findings can be generalized to other settings or groups.

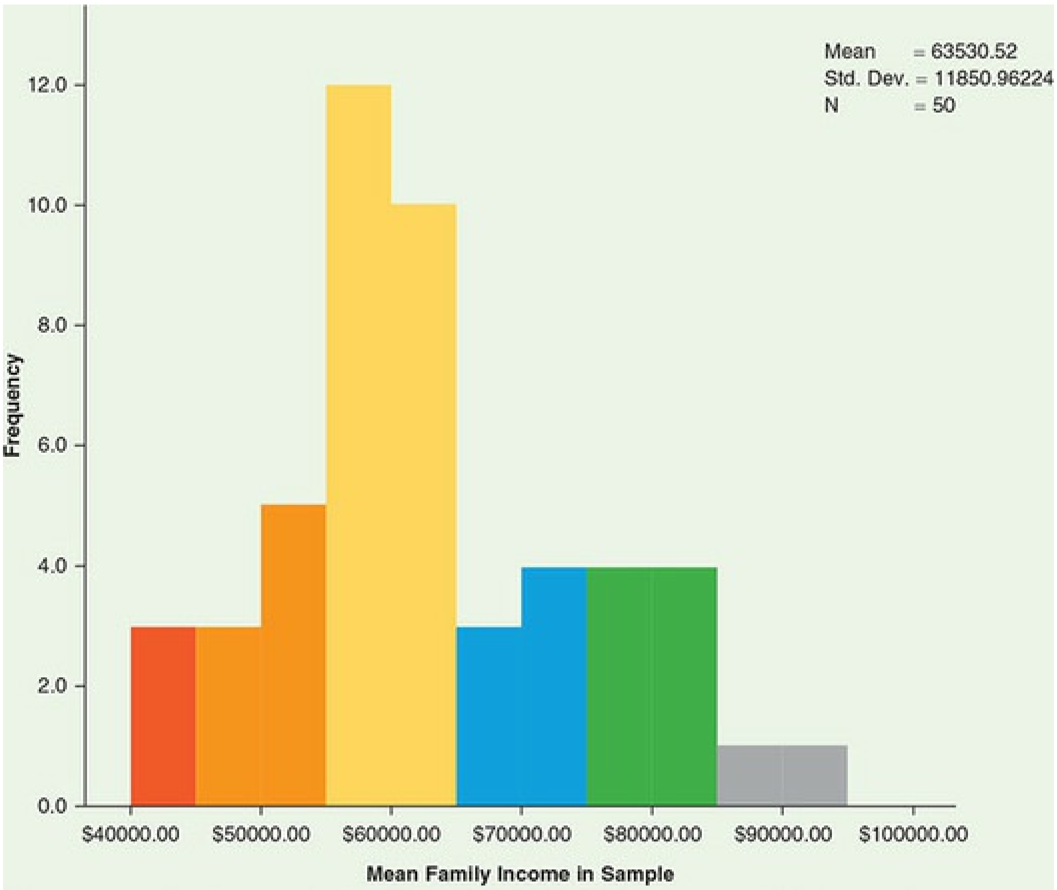
Sampling Distributions

A well-designed probability sample is one that is likely to be representative of the population from which it was selected. But as you've seen, random samples still are subject to sampling error owing just to chance. To deal with that problem, social researchers consider the properties of a sampling distribution, a hypothetical distribution of a statistic across all the random samples that could be drawn from a population. Any single random sample can be thought of as just one of an infinite number of random samples that, in theory, could have been selected from the population. If we had the finances of Gatsby and the patience of Job and were able to draw an infinite number of samples, and we calculated the same type of statistic for each of these samples, we would then have a sampling distribution. Understanding sampling distributions is the foundation for understanding how statisticians can estimate sampling error.

What does a sampling distribution look like? Because a sampling distribution is based on some statistic calculated for different samples, we need to choose a statistic. Let's focus on the arithmetic average, or mean. I will explain the calculation of the mean in Chapter 9, but you may already be familiar with it: You add up the values of all the cases and divide by the total number of cases. Let's say you draw a random sample of 500 families and find that their average (mean) family income is $58,239. Imagine that you then draw another random sample. That sample's mean family income might be $60,302. Imagine marking these two means on graph paper and then drawing more random samples and marking their means on the graph. The resulting graph would be a sampling distribution of the mean.

Exhibit 5.16 demonstrates what happened when I did something very similar to what I have just described—not with an infinite number of samples and not from a large population, but through the same process using the 2012 General Social Survey (GSS) sample as if it were a population. First, I drew 50 different random samples, each consisting of 30 cases, from the 2012 GSS. (The standard notation for the number of cases in each sample is $n = 30$.) Then I calculated for each random sample the approximate mean family income (approximate because the GSS does not record actual income in dollars). I then graphed the means of the 50 samples. Each bar in Exhibit 5.16 shows how many samples had a mean family income in each $5,000 category between $40,000 and $95,000. The mean for the "population" (the total GSS sample in this example) is $64,238, and you can see that many of the samples in the sampling distribution are close to this value, with the mean for this sampling distribution being $63,530.52—almost identical to the "population" mean. However, although many of the sample means are close to the population mean, some are quite far from it (the lowest is actually $43,508, while the highest is $92,930). If you had calculated the mean from only one sample, it could have been anywhere in this sampling distribution, but it is *unlikely* to have been far from the population mean—that is, unlikely to have been close to either end (or “tail”) of the distribution.

Exhibit 5.16 Partial Sampling Distribution: Mean Family Income (Samples of Size 30)*



*Mean income in highest recorded category (≥ $150,000) set as 246,484 based on distribution in Bureau of Census, Current Population Survey, Family Income Table, 2012, FINC-07. http://www.census.gov/hhes/www/cpstables/032013/faminc/toc.htm

Output generated from IBM SPSS Statistics, Version 21.

Source: General Social Survey 2012 (National Opinion Research Center [NORC] 2014).

Estimating Sampling Error

We don't actually observe sampling distributions in real research; researchers just draw the best sample they can and then are stuck with the results—one sample, not a distribution of samples. A sampling distribution is a theoretical distribution. However, we can use the properties of sampling distributions to calculate the amount of sampling error that was likely with the random sample used in a study. The tool for calculating sampling error is called inferential statistics.

Sampling distributions for many statistics, including the mean, have a "normal" shape. A graph of a normal distribution looks like a bell, with one hump in the middle, centered on the population mean, and the number of cases tapering off on both sides of the mean. Note that a normal distribution is symmetric: If you folded it in half at its center (at the population mean), the two halves would match perfectly. This shape is produced by random sampling error—variation owing purely to chance. The value of the statistic varies from sample to sample because of chance, so higher and lower values are equally likely.

The partial sampling distribution in Exhibit 5.16 does not have a completely normal shape because it involves only a small number of samples (50), each of which has only 30 cases. Exhibit 5.17 shows what the sampling distribution of family incomes would look like if it formed a perfectly normal distribution—if, rather than 50 random samples, I had selected thousands of random samples.

The properties of a sampling distribution facilitate the process of statistical inference. In the sampling distribution, the most frequent value of the sample statistic—the statistic (such as the mean) computed from sample data—is identical to the population parameter—the statistic computed for the entire population. In other words, we can have a lot of confidence that the value at the peak of the bell curve represents the norm for the entire population. A population parameter also may be termed the *true value* for the statistic in that population. A sample statistic is an estimate of a population parameter.

In a normal distribution, a predictable proportion of cases fall within certain ranges. Inferential statistics takes advantage of this feature and allows researchers to estimate how likely it is that, given a particular sample, the true population value will be within some range of the statistic. For example, a statistician might conclude from a sample of 30 families, "We can be 95% confident that the true mean family income in the total population is between $39,037 and $89,977." The interval from $39,037 to $89,977 would then be called the *95% confidence interval for the mean.* The lower ($39,037) and upper ($89,977) bounds of this interval are termed the *confidence limits.* Exhibit 5.17 marks such confidence limits, indicating the range that encompasses 95% of the area under the normal curve; 95% of all sample means would fall within this range, as does the mean of our hypothetical sample of 30 cases.

Inferential statistics: A mathematical tool for estimating how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

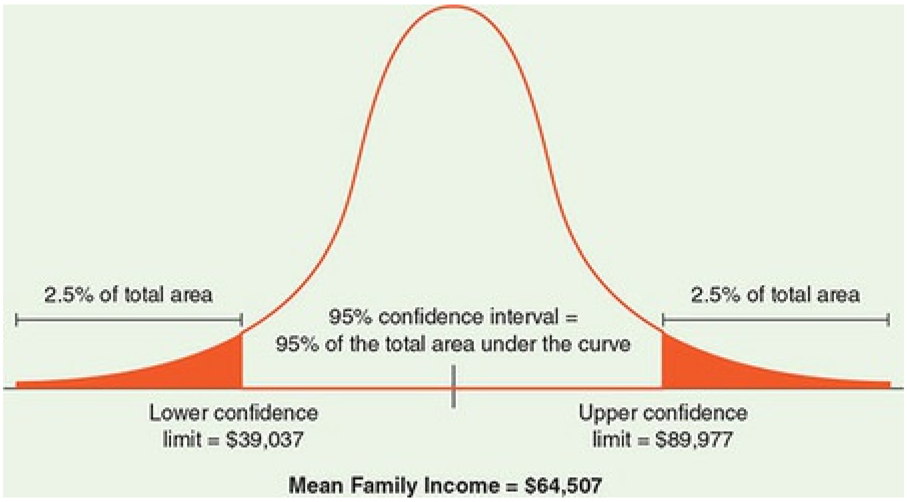
Random sampling error (chance sampling error): Differences between the population and the sample that are due only to chance factors (random error), not to systematic sampling error. Random sampling error may or may not result in an unrepresentative sample. The magnitude of sampling error resulting from chance factors can be estimated statistically.

Sample statistic: The value of a statistic, such as a mean, computed from sample data.

Population parameter: The value of a statistic, such as a mean, computed using the data for the entire population; a sample statistic is an estimate of a population parameter.

Although all normal distributions have these same basic features, they differ from one another in the extent to which they cluster around the mean. A sampling distribution is more compact when it is based on larger samples. Stated another way, we can be more confident in estimates based on larger random samples because we know that a larger sample creates a more compact sampling distribution. Compare the two sampling distributions of mean family income shown in Exhibit 5.18. Both depict the results for about 50 samples. In one study, each sample consisted of 100 families, but in the other study, each sample consisted of only 5 families. Clearly, the larger samples result in a sampling distribution that is much more tightly clustered around the mean (range of \$53,109 to \$83,563) than is the case with the smaller samples (range of \$17,100 to \$117,297). The 95% confidence interval for mean family income for the entire 2012 GSS sample of 1,758 cases (the ones that had valid values of family income) was \$61,243 to \$67,233—an interval only \$5,990 wide. But the 95% confidence interval for the mean family income in one GSS subsample of 100 cases was much wider: \$23,601. And for one of the subsamples of only 5 cases, the 95% confidence interval was very broad indeed: \$177,430. As you can see, such small samples result in statistics that actually give us very little useful information about the population.

Exhibit 5.17 Normal Sampling Distribution: Mean Family Income



Other confidence intervals, such as the 99% confidence interval, can be reported. As a matter of convention, statisticians use only the 95%, 99%, and 99.9% confidence limits to estimate the range of values that are likely to contain the true value. These conventional limits reflect the conservatism inherent in classical statistical inference: Don't make an inferential statement unless you are very confident (at least 95% confident) that it is correct.

The less precise an estimate of a particular statistic from a particular sample is, the more confident we can be—and the wider the confidence interval will be. As I mentioned previously, the 95% confidence interval of family income (in approximate dollars) for the entire 2012 GSS sample is $61,243 to $67,233 (a width of $5,990); the 99% confidence interval is $60,326 to $68,150 (a width of $7,824).

I will explain how to calculate confidence intervals in Chapter 9 and how to express the variability in a sample estimate with a statistic called the *standard error.* The basic statistics that I introduce in that chapter will make it easier to understand these other statistics. If you have already completed a statistics course, you might want to turn now to Chapter 9's confidence interval section for a quick review. In any case, you should now have a sense of how researchers make inferences from a random sample of a population.

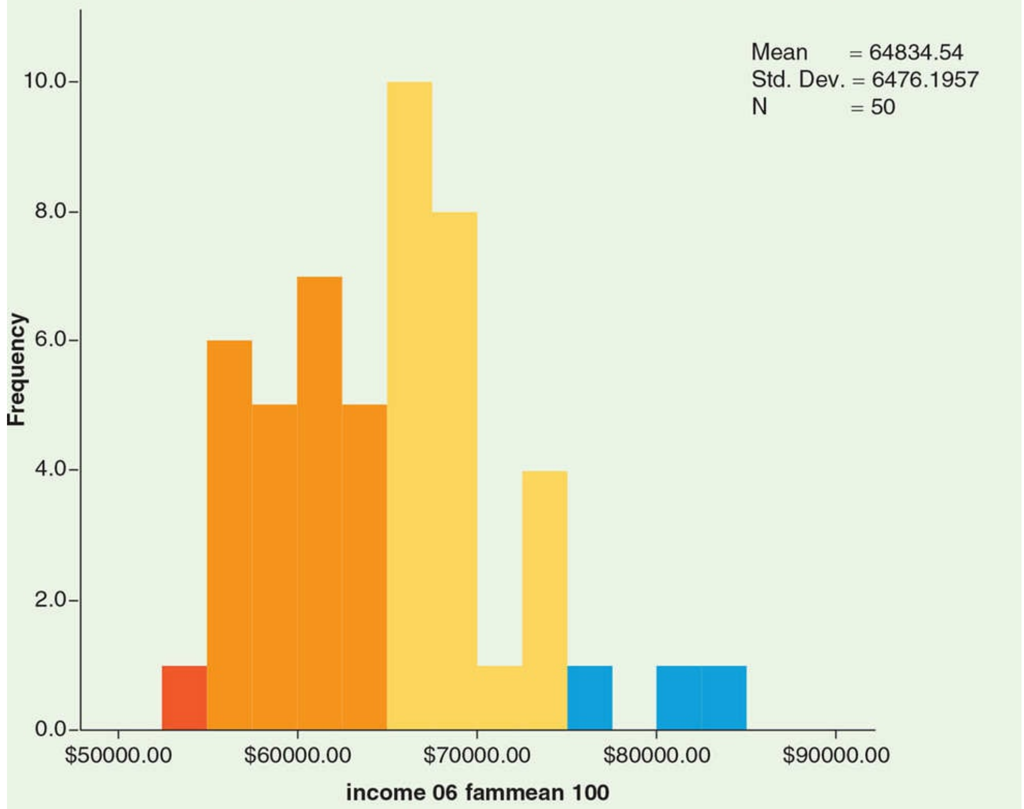
Sample Size Considerations

You have learned now that more confidence can be placed in the generalizability of statistics from larger samples, so that you may be eager to work with random samples that are as large as possible. Unfortunately, researchers often cannot afford to sample a very large number of cases. Therefore, they try to determine during the design phase of their study how large a sample they must have to achieve their purposes. They have to consider the degree of confidence desired, the homogeneity of the population, the complexity of the analysis they plan, and the expected strength of the relationships they will measure.

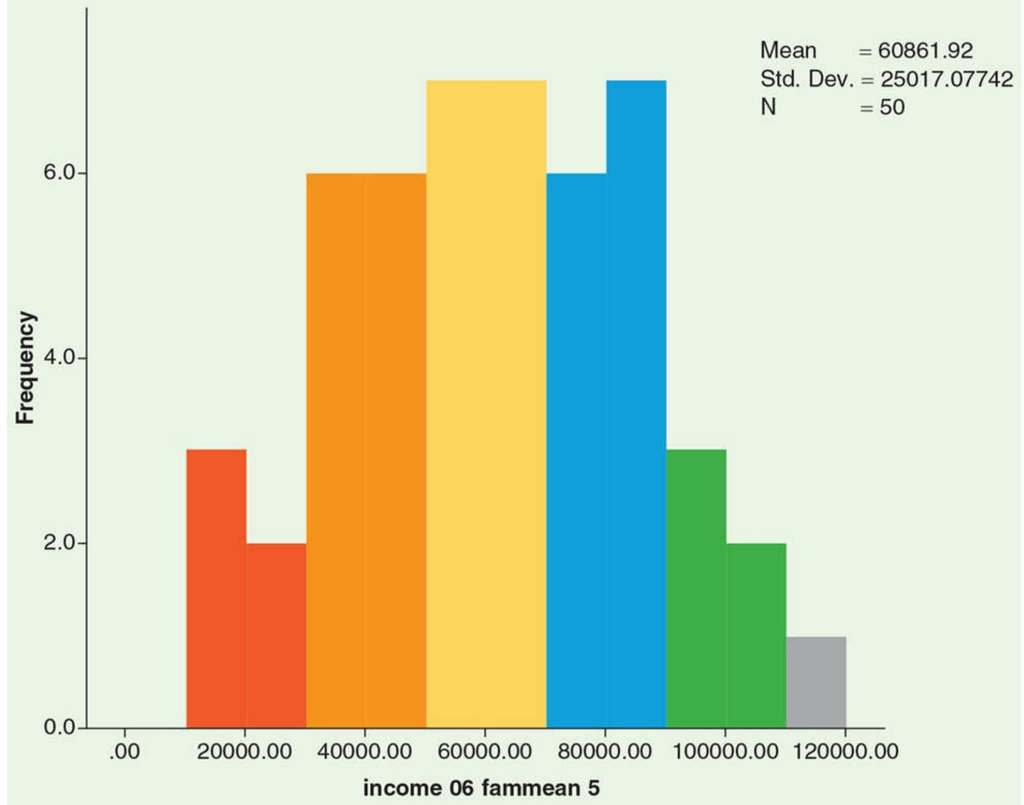
- The less sampling error desired, the larger the sample size must be.
- Samples of more homogeneous populations can be smaller than can samples of more diverse populations. Stratified sampling uses prior information on the population to create more homogeneous population strata from which the sample can be selected, so stratified samples can be smaller than simple random samples.
- If the only analysis planned for a survey sample is to describe the population in terms of a few variables, a smaller sample is required than if a more complex analysis involving sample subgroups is planned. If much of the analysis will focus on estimating the characteristics of subgroups within the sample, the size of the subgroups must be considered, rather than the size of the total sample (Levy and Lemeshow 1999:74).

Exhibit 5.18a The Effect of Sample Size on Sampling Distributions

Partial Sampling Distribution: Mean Family Income (Samples of Size 100)*



Partial Sampling Distribution: Mean Family Income (Samples of Size 5)*



*Mean income in highest recorded category (≥ $150,000) set as 246,484 based on distribution in Bureau of Census, Current Population Survey, Family Income Table, 2012, FINC-07. http://www.census.gov/hhes/www/cpstables/032013/faminc/toc.htm

Output generated from IBM SPSS Statistics, Version 21.

Source: General Social Survey 2012 (National Opinion Research Center [NORC] 2014).

- When the researchers expect to find very strong relationships between the variables when they test hypotheses, they will need a smaller sample to detect these relationships than if they expect weaker relationships.

Researchers can make more precise estimates of the sample size required through a method called *statistical power analysis* (Kraemer and Thiemann 1987). Statistical power analysis requires a good advance estimate of the strength of the hypothesized relationship in the population. In addition, the math is complicated, so it helps to have some background in mathematics or to be able to consult a statistician. For these reasons, many researchers do not conduct formal power analyses when deciding how many cases to sample.

You can obtain some general guidance about sample sizes from the current practices of social scientists. For professional studies of the national population in which only a simple description is desired, professional social science studies typically have used a sample size of between 1,000 and 1,500 people, with as many as 2,500 being included if detailed analyses are planned. Studies of local or regional populations often sample only a few hundred people, partly because these studies lack sufficient funding to draw larger samples. Of course, the sampling error in these smaller studies is considerably larger than in a typical national study (Sudman 1976:87).

Conclusions

Sampling is a powerful tool for social science research. Probability sampling methods allow a researcher to use the laws of chance, or probability, to draw samples from which population parameters can be estimated with a high degree of confidence. A sample of just 1,000 or 1,500 individuals can be used to estimate reliably the characteristics of the population of a nation comprising millions of individuals.

But researchers do not come by representative samples easily. Well-designed samples require careful planning, some advance knowledge about the population to be sampled, and adherence to systematic selection procedures—all so that the selection procedures are not biased. And even after the sample data are collected, the researcher's ability to generalize from the sample findings to the population is not completely certain. The best that he or she can do is to perform additional calculations that state the degree of confidence that can be placed in the sample statistic.

The alternatives to random, or probability-based, sampling methods are almost always much less palatable for quantitative studies, even though they are typically much cheaper. Without a method of selecting cases likely to represent the population in which the researcher is interested, research findings will have to be carefully qualified. Qualitative researchers whose goal is to understand a small group or setting in depth may necessarily have to use unrepresentative samples, but they must keep in mind that the generalizability of their findings will not be known. Additional procedures for sampling in qualitative studies are introduced in Chapter 10.

Social scientists often seek to generalize their conclusions from the population that they studied to some larger target population. The validity of generalizations of this type is necessarily uncertain, because having a representative sample of a particular population does not at all ensure that what we find will hold true in other populations. Nonetheless, as you will see in Chapter 16, the cumulative findings from studies based on local or otherwise unrepresentative populations can provide important information about broader populations.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Availability sampling 169
Census 154
Cluster 165
Disproportionate stratified sampling 163
Elements 149
Enumeration units 150
Inferential statistics 178
Multistage cluster sampling 165
Nonprobability sampling method 156
Nonrespondents 157
Periodicity 162
Population 149
Population parameter 178
Probability of selection 156
Probability sampling method 156
Proportionate stratified sampling 163
Purposive sampling 172
Quota sampling 170
Random digit dialing 160
Random number table 160
Random sampling 156
Random sampling error (chance sampling error) 178
Replacement sampling 160
Representative sample 154
Sample 149
Sample statistic 178
Sampling error 152
Sampling frame 149
Sampling interval 162
Sampling units 150
Simple random sampling 160
Snowball sampling 173
Stratified random sampling 162
Systematic bias 157
Systematic random sampling 162
Target population 152

Highlights

- Sampling theory focuses on the generalizability of descriptive findings to the population from which the sample was drawn. It also considers whether statements can be generalized from one population to another.
- Sampling is unnecessary when the elements that would be sampled are identical, but the complexity of the social world makes it difficult to argue very often that all different elements are identical. Conducting a complete census of a population also eliminates the need for sampling, but the resources required for a complete census of a large population are usually prohibitive.
- Nonresponse undermines sample quality: The obtained sample, not the desired sample, determines sample quality.
- Probability sampling methods rely on a random selection procedure to ensure there is no systematic bias in the selection of elements. In a probability sample, the odds of selecting elements are known, and the method of selection is carefully controlled.
- A sampling frame (a list of elements in the population) is required in most probability sampling methods. The adequacy of the sampling frame is an important determinant of sample quality.
- Simple random sampling and systematic random sampling are equivalent probability sampling methods in most of the situations. However, systematic random sampling is inappropriate for sampling from lists of elements that have a regular, periodic structure.
- Stratified random sampling uses prior information about a population to make sampling more efficient. Stratified sampling may be either proportionate or disproportionate. Disproportionate stratified sampling is useful when a research question focuses on a stratum or on strata that make up a small proportion of the population.
- Multistage cluster sampling is less efficient than simple random sampling, but it is useful when a sampling frame is unavailable. It is also useful for large populations spread out across a wide area or among many organizations.
- Nonprobability sampling methods can be useful when random sampling is not possible, when a research question does not concern a larger population, and when a preliminary exploratory study is appropriate. However, the representativeness of nonprobability samples cannot be determined.
- The likely degree of error in an estimate of a population characteristic based on a probability sample decreases when the size of the sample and the homogeneity of the population from which the sample was selected increase. The proportion of the population that is sampled does not affect sampling error, except when that proportion is large. The degree of sampling error affecting a sample statistic can be estimated from the characteristics of the sample and knowledge of the properties of sampling distributions.

Discussion Questions

1. When (if ever) is it reasonable to assume that a sample is not needed because “everyone is the same”—that is, the population is homogeneous? Does this apply to research such as that of Stanley Milgram on obedience to authority? What about investigations of student substance abuse? How about investigations of how people (or their bodies) react to alcohol? What about research on likelihood of voting (the focus of Chapter 9)?
2. All adult U.S. citizens are required to participate in the decennial census, but some do not. Some social scientists have argued for putting more resources into a large representative sample, so that more resources are available to secure higher rates of response from hard-to-include groups. Do you think that the U.S. Census should shift to a probability-based sampling design? Why or why not?
3. What increases sampling error in probability-based sampling designs? Stratified rather than simple random sampling? Disproportionate (rather than proportionate) stratified random sampling? Stratified rather than cluster random sampling? Why do researchers select *disproportionate* (rather than proportionate) stratified samples? Why do they select cluster rather than simple random samples?
4. What are the advantages and disadvantages of probability-based sampling designs compared with nonprobability-based designs? Could any of the research described in this chapter with a nonprobability-based design have been conducted instead with a probability-based design? What are the difficulties that might have been encountered in an attempt to use random selection? How would you discuss the degree of confidence you can place in the results obtained from research using a nonprobability-based sampling design?

Practice Exercises

1. Select a random sample using the table of random numbers in Appendix C. Compute a statistic based on your sample and compare it with the corresponding figure for the entire population. Here's how to proceed:
    a. Select a very small population for which you have a reasonably complete sampling frame. One possibility would be the list of asking prices for houses advertised in your local paper. Another would be the listing of some characteristic of states in a U.S. Census Bureau publication, such as average income or population size.
    b. Create your sampling frame, a numbered list of all the elements in the population. If you are using a complete listing of all elements, as from a U.S. Census Bureau publication, the sampling frame is the same as the list. Just number the elements (states). If your population is composed of housing ads in the local paper, your sampling frame will be those ads that contain a housing price. Identify these ads, and then number them sequentially, starting with 1.
    c. Decide on a method of picking numbers out of the random number table in Appendix C, such as taking every number in each row, row by row (or you may move down or diagonally across the columns). Use only the first (or last) digit in each number if you need to select 1 to 9 cases, or only the first (or last) two digits if you want fewer than 100 cases.
    d. Pick a starting location in the random number table. It's important to pick a starting point in an unbiased way, perhaps by closing your eyes and then pointing to some part of the page.
    e. Record the numbers you encounter as you move from the starting location in the direction you decided on in advance, until you have recorded as many random numbers as the number of cases you need in the sample. If you are selecting states, 10 might be a good number. Ignore numbers that are too large (or too small) for the range of numbers used to identify the elements in the population. Discard duplicate numbers.
    f. Calculate the average value in your sample for some variable that was measured—for example, population size in a sample of states or housing price for the housing ads. Calculate the average by adding the values of all the elements in the sample and dividing by the number of elements in the sample.
    g. Go back to the sampling frame and calculate this same average for all elements in the list. How close is the sample average to the population average?
    h. Estimate the range of sample averages that would be likely to include 90% of the possible samples.
2. Draw a snowball sample of people who are involved in bungee jumping or some other uncommon sport that does not involve teams. Ask friends and relatives to locate a first contact, and then call or visit this person and ask for names of others. Stop when you have identified a sample of 10. Review the problems you encountered, and consider how you would proceed if you had to draw a larger sample.
3. Two lesson sets from the "Interactive Exercises" link on the study site will help you review terminology, "Identifying Sampling Techniques" and the logic of "Assessing Generalizability."
4. Identify one article at the book's study site that used a survey research design. Describe the sampling procedure. What type was it? Why did the author(s) use this particular type of sample?

Ethics Questions

1. How much pressure is too much pressure to participate in a probability-based sample survey? Is it okay for the U.S. government to mandate legally that all citizens participate in the decennial census? Should companies be able to require employees to participate in survey research about work-related issues? Should students be required to participate in surveys about teacher performance? Should parents be required to consent to the participation of their high school–age students in a survey about substance abuse and health issues? Is it okay to give monetary incentives for participation in a survey of homeless shelter clients? Can monetary incentives be coercive? Explain your decisions.
2. Federal regulations require special safeguards for research on persons with impaired cognitive capacity. Special safeguards are also required for research on prisoners and on children. Do you think special safeguards are necessary? Why or why not? Do you think it is possible for individuals in any of these groups to give "voluntary consent" to research participation? What procedures might help make consent to research truly voluntary in these situations? How could these procedures influence sampling plans and results?

Web Exercises

1. Research on homelessness has been rising in recent years as housing affordability has declined. Search the web for sites that include the word *homelessness* and see what you find. You might try limiting your search to those that also contain the word *census.* Pick a site and write a paragraph about what you learned from it.
2. Check out the "Income & Poverty" section of the U.S. Census Bureau website (https://www.census.gov/topics/income-poverty.html). Based on the data you find there, write a brief summary of some aspects of the current characteristics of the U.S. population.

Video Interview Questions

Listen to the researcher interview for Chapter 5 at edge.sagepub.com/schutt9e.

1. What was Anthony Roman's research question in his phone survey research study?
2. What were Roman's major discoveries in this project? How does this emphasize the importance of sampling selectively and carefully?

SPSS Exercises

1. Take a look again at the distribution of support for capital punishment (CAPPUN), this time with what is called a *frequency distribution.*
    a. Click Analyze/Descriptive Statistics/Frequencies.
    b. Highlight CAPPUN and click on the arrow that sends it over to the Variables window, then click OK.

    Examine the percentages in the Valid percent column. What percentage of the U.S. population in 2016 favored capital punishment?
2. Now select random samples of the GSS2016 or GSS2016x respondents (sorry, but you can't carry out this exercise if you are using the GSS2016x_reduced file):
    a. Go to the Data Editor window, and select a random sample containing 40 of the respondents. From the menu:
    b. Click Data/Select cases/Random sample of cases.
    c. Choose *S*ample.
    d. For "Sample Size," choose Exactly and then enter *40* cases from the first *100* cases.
    e. Click Continue/OK. (Before you click OK, be sure that the "Filter out unselected cases" box is checked.)
    f. Determine the percentage of the subsample that favored capital punishment by repeating the steps in SPSS Exercise 1. Record the subsample characteristics and its percentage.
    g. Now, repeat Steps 2a through 2e 10 times. Each time, add 100 to the "first 100 cases" request (so that on the last step you will be requesting "Exactly 40 cases from the first 1,000 cases").
    h. Select a random sample containing five of the respondents. Now repeat Steps 2a through 2f (10 times), this time for samples of 5.
3. How does the distribution of CAPPUN in these subsamples compare with that for the total GSS sample?
    a. Plot the percents produced in Step 2g and 2h on separate sheets of graph paper. Each graph's horizontal axis will represent the possible range of percentages (from 0 to 100, perhaps in increments of 5); the vertical axis will represent the number of samples in each range of percentages (perhaps ranging from 0 to 10). Make an X to indicate this percentage for each sample. If two samples have the same percentage, place the corresponding Xs on top of each other. The X for each sample should be one unit high on the vertical axis.
    b. Draw a vertical line corresponding to the point on the horizontal axis that indicates the percentage of the total GSS sample that favors capital punishment.
    c. Describe the shape of both graphs. These are the sampling distributions for the two sets of samples. Compare them with each other. Do the percentages from the larger samples tend to be closer to the mean of the entire sample (as obtained in SPSS Exercise 1)? What does this tell you about the relationship between sample size and sampling error?

Developing a Research Proposal

Consider the possibilities for sampling (Exhibit 3.10, #8).

1. Propose a sampling design that would be appropriate if you were to survey students on your campus only. Define the population, identify the sampling frame(s), and specify the elements and any other units at different stages. Indicate the exact procedure for selecting people to be included in the sample.
2. Propose a different sampling design for conducting your survey in a larger population, such as your city, state, or the entire nation.

Chapter 6 Research Design and Causation

Research That Matters, Questions That Count
Research Design Alternatives
  Units of Analysis
    Individual and Group
Research in the News: Police and Black Drivers
  The Ecological Fallacy and Reductionism
  Cross-Sectional and Longitudinal Designs
  Cross-Sectional Designs
  Longitudinal Designs
Quantitative or Qualitative Causal Explanations
  Quantitative (Nomothetic) Causal Explanations
  Qualitative (Idiographic) Causal Explanations
Careers and Research
Criteria and Cautions for Nomothetic Causal Explanations
  Association
  Time Order
    Experimental Designs
    Nonexperimental Designs
  Nonspuriousness
    Randomization
    Statistical Control
  Mechanism
  Context
Comparing Research Designs
Conclusions

Identifying causes—figuring out why things happen—is the goal of most social science research. Unfortunately, valid explanations of the causes of social phenomena do not come easily. Why did the number and rate of homicides rise in the early 1990s and then begin a sustained drop that has continued in the 2000s, even during the 2008–2010 recession, to a level last seen in 1968 (Smith and Cooper 2013) (see Exhibit 6.1)? Arizona State University criminologist Scott Decker points to the low levels of crime committed by illegal immigrants to explain the falling crime rate in his state (Archibold 2010), whereas criminal justice advocates in Texas point to the state's investment in community treatment and diversion programs (Grissom 2011). Police officials in New York City point to the effectiveness of CompStat, a computer program that highlights the location of crimes (Dewan 2004a:A25; Dewan 2004b:A1; Kaplan 2002:A3), but others think New York City has benefited from its Safe Streets, Safe Cities program (Rashbaum 2002) or from a decline in crack cocaine use (Dewan 2004b:C16). Should we worry about the consequences for crime of the increasing number of drug arrests nationally (Bureau of Justice Statistics 2011) and a rise in abuse of prescription drugs (Goodnough 2010)? Now we know from the Desmond, Papachristos, and Kirk (2013) study that we also need to take account of factors that may decrease willingness to report crimes to the police. What features of research designs can help us answer questions like these?

Research That Matters, Questions That Count

Frank Jude and Lovell Harris left a party hosted by a Milwaukee police officer in a middle-class white neighborhood on the night of October 23, 2004. They had been invited to the party by friends of the white officer, but left after feeling uncomfortable, only to be confronted outside by a group of 10 men. Jude and Harris, both black, were attacked; Harris escaped but Jude was beaten severely and hospitalized. The incident was not widely publicized until three months later, but at that point protesters mobilized and subsequently nine officers were fired and four others disciplined. After extensive litigation, more protests, and a federal investigation, seven officers were convicted in federal court. As in too many such incidents, concerns were expressed about the potential effect on residents of black neighborhoods reporting crime to the police.

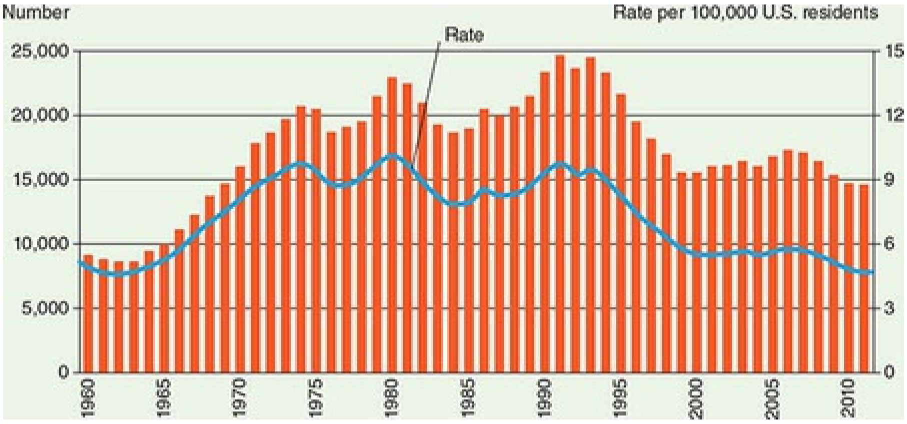
Sociologists Matthew Desmond (Harvard), Andrew Papachristos (Yale), and David Kirk (Oxford) decided to investigate the effect of the incident on crime reporting. They were familiar with prior research on the extent of "legal cynicism" in many poor, minority communities, but they also knew that prior research suggested that blacks and whites were equally likely to say they would report crime to the police (Desmond, Papachristos, and Kirk 2016). It was not so obvious that an incident like Jude's publicized beating would cause a change in reporting behavior. Desmond, Papachristos, and Kirk analyzed all 883,146 emergency (911) calls reporting crime that were placed in Milwaukee between March 1, 2004, and December 31, 2010. The location of these calls allowed the researchers to determine the racial percentages in the block group where the call originated. They found that after reports of Jude's beating appeared in the press, Milwaukee residents, particularly those living in black neighborhoods, were less likely to call the police—even to report violent crime.

1. What effects would you hypothesize are caused by such publicized incidents?
2. Based on this brief description, how strong do you think the evidence is for a conclusion that Frank Jude's beating caused a decline in crime reporting to the police in black Milwaukee neighborhoods?
3. How would you suggest designing a study to test the hypothesis that "publicized cases of apparent police violence against unarmed black men" cause a decline in 911 calls about crime?

In this chapter, you will learn about the implications of research design features for causal analyses. By the end of the chapter, you will understand why testing causal hypotheses can be difficult and how to design research to strengthen causal conclusions. As you read the chapter, extend you understanding by reading the 2016 *Youth Justice* article by Desmond, Papachristos, and Kirk at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 6 at edge.sagepub.com/schutt9e.

Desmond, Matthew, Andrew Papachristos, and David Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81(5):857–876.

Exhibit 6.1 Number and Rate of Homicides in the United States, 1960–2011



Source: Federal Bureau of Investigation. 2013. “Table 1: Crime in the United States.” Uniform Crime Reports.

Research Design Alternatives

I begin this chapter by discussing three key elements of research design: the design's units of analysis, its use of cross-sectional or longitudinal data, and whether its methods are primarily quantitative or qualitative. Whenever we design research, we must decide whether to use individuals or groups as our units of analysis and whether to collect data at one or several points in time. The decisions that we make about these design elements will affect our ability to draw causal conclusions in our analysis. Whether the design is primarily quantitative or qualitative in its methods also affects the type of causal explanation that can be developed: Quantitative projects lead to *nomothetic* causal explanations, whereas qualitative projects that have a causal focus can lead to *idiographic* explanations (described in the section "Quantitative or Qualitative Causal Explanations"). After reviewing these three key design elements, I will also review the criteria for achieving explanations that are causally valid from a nomothetic perspective. By the end of the chapter, you should have a good grasp of the different meanings of causation and be able to ask the right questions to determine whether causal inferences are likely to be valid. You also may have a better answer about the causes of crime and violence.

Units of Analysis

In nonexperimental research designs, we can be misled about the existence of an association between two variables when we do not know to what units of analysis the measures in our study refer—that is, the level of social life on which the research question is focused, such as individuals, groups, towns, or nations. I first discuss this important concept before explaining how it can affect causal conclusions.

Units of analysis: The level of social life on which a research question is focused, such as individuals, groups, towns, or nations.

Individual and Group

In most sociological and psychological studies, the units of analysis are individuals. The researcher may collect survey data from individuals, analyze the data, and then report on, say, how many individuals felt socially isolated and whether substance abuse by individuals was related to their feelings of social isolation.

The units of analysis may instead be groups of some sort, such as families, schools, work organizations, towns, states, or countries. For example, a researcher may collect data from town and police records on the number of accidents in which a driver was intoxicated and the presence or absence of a server liability law in the town. (These laws make those who serve liquor liable for accidents caused by those to whom they served liquor.) The researcher can then analyze the relationship between server liability laws and the frequency of accidents caused by drunk driving (perhaps also taking into account the town's population). Because the data describe the town, towns are the units of analysis.

In some studies, groups are the units of analysis, but data are collected from individuals. For example, in their study of influences on violent crime in Chicago neighborhoods, Robert Sampson, Stephen Raudenbush, and Felton Earls (1997:919) hypothesized that efficacy would influence neighborhood crime rates. *Collective efficacy* was defined conceptually as a characteristic of the neighborhood: the extent to which residents were likely to help other residents and were trusted by other residents. However, Sampson et al. measured this variable in a survey of individuals. The responses of individual residents about their perceptions of their neighbors' helpfulness and trustworthiness were averaged together to create a collective efficacy score for each neighborhood. This neighborhood measure of collective efficacy was used to explain variation in the rate of violent crime between neighborhoods. The data were collected from individuals and were about individuals, but they were combined (aggregated) to describe neighborhoods. The units of analysis were thus groups (neighborhoods).

In a study such as that of Sampson et al. (1997), we can distinguish the concept of units of analysis from the units of observation. Data were collected from individuals, the units of observation in this study, and then the data were aggregated and analyzed at the group level. In most studies, the units of observation and the units of analysis are the same. For example, Yili Xu, Mora Fiedler, and Karl Flaming (2005), in collaboration with the Colorado Springs Police Department, surveyed a stratified random sample of 904 residents to test whether their sense of collective efficacy and other characteristics would predict their perceptions of crime, fear of crime, and satisfaction with police. Their data were collected from individuals and analyzed at the individual level. They concluded that collective efficacy was not as important as in Sampson et al.'s (1997) study.

The important point is to know what the units of observation are and what the level of analysis is, and then to evaluate whether the conclusions are appropriate to these study features. A conclusion that "crime increases with joblessness" could imply either that individuals who lose their jobs are more likely to commit a crime or that a community with a high unemployment rate is likely to have a high crime rate—or both. Whether we are drawing conclusions from data we collected or interpreting others' conclusions, it is important to be clear about the relationship to which the results refer.

We also have to know what the units of analysis are to interpret statistics appropriately. Measures of association tend to be stronger for group-level than for individual-level data because measurement errors at the individual level tend to cancel out at the group level (Bridges and Weis 1989:29–31).

The Ecological Fallacy and Reductionism

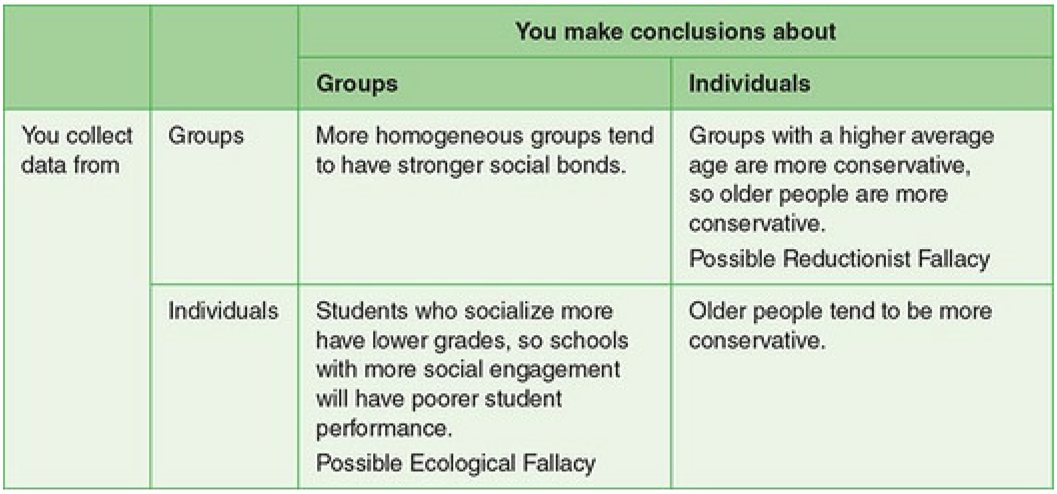
Researchers should make sure that their causal conclusions reflect the units of analysis in their study. Conclusions about processes at the individual level should be based on individual-level data; conclusions about group-level processes should be based on data collected about groups. When this rule is violated, you can often be misled about the existence of an association between two variables.

Units of observation: The cases about which measures actually are obtained in a sample.

A researcher who draws conclusions about individual-level processes from group-level data could be making what is termed an ecological fallacy (see Exhibit 6.2). The conclusions may or may not be correct, but we must recognize the fact that group-level data do not necessarily reflect solely individual-level processes. For example, a researcher may examine factory records and find that the higher the percentage of unskilled workers in factories, the higher the rate of employee sabotage in those factories. But the researcher would commit an ecological fallacy if he or she then concluded that individual unskilled factory workers are more likely to engage in sabotage. This conclusion is about an individual-level causal process (the relationship between the occupation and criminal propensities of individuals), even though the data describe groups (factories). It could actually be that white-collar workers are the ones more likely to commit sabotage in factories with more unskilled workers, perhaps because the white-collar workers feel they won't be suspected in these settings.

Ecological fallacy: An error in reasoning in which incorrect conclusions about individual-level processes are drawn from group-level data.

Exhibit 6.2 Errors in Causal Conclusions



		You make conclusions about	
		Groups	**Individuals**
You collect data from	Groups	More homogeneous groups tend to have stronger social bonds.	Groups with a higher average age are more conservative, so older people are more conservative. Possible Reductionist Fallacy
	Individuals	Students who socialize more have lower grades, so schools with more social engagement will have poorer student performance. Possible Ecological Fallacy	Older people tend to be more conservative.

Conversely, when data about individuals are used to make inferences about group-level processes, a problem occurs that can be thought of as the mirror image of the ecological fallacy: the reductionist fallacy, also known as *reductionism,* or the *individualist fallacy* (see Exhibit 6.2). For example, a reductionist explanation of individual violence would focus on biological factors, such as genes or hormones, rather than on the community's level of social control. Similarly, a reductionist explanation of behavior problems in grade school classrooms would focus on the children's personalities, rather than on classroom structure, teacher behavior, or the surrounding neighborhood.

There is, of course, nothing inherently wrong with considering biological, psychological, or other factors that affect individuals when explaining behavior in social contexts (Sayer 2003). The key issue is whether there are properties of groups or other aggregates that are more than "the sum of their parts." If there are such "higher level properties," we can speak of "emergence" at the group level and we need to be careful to avoid a reductionist fallacy.

Reductionist fallacy (reductionism): An error in reasoning that occurs when incorrect conclusions about group-level processes are based on individual-level data; also known as an individualist fallacy.

Emergence: The appearance of phenomena at a group level that cannot be explained by the properties of individuals within the group; emergence implies phenomena that are more than “the sum of their parts.”

The fact that errors in causal reasoning can be made in this way should not deter you from conducting research with group-level data nor make you unduly critical of researchers who make inferences about individuals on the basis of group-level data. When considered broadly, many research questions point to relationships that could be manifested in many ways and on many levels. Sampson’s (1987) study of urban violence is a case in point. His analysis involved only aggregate data about cities, and he explained his research approach as partly a response to the failure of other researchers to examine this problem at the structural, aggregate level. Moreover, Sampson argued that the rates of joblessness and family disruption in communities influence community social processes, not just individuals who are unemployed or who grew up without two parents. Yet Sampson suggested that the experience of joblessness and poverty is what tends to reduce the propensity of individual men to marry and that the experience of growing up in a home without two parents, in turn, increases the propensity of individual juveniles to commit crimes. These conclusions about individual behavior seem consistent with the patterns Sampson found in his aggregate, city-level data, so it seems unlikely that he committed an ecological fallacy when he proposed them.

In the News Research in the News: Police and Black Drivers

For Further Thought?

Social science researchers at Stanford University examined indications of racial disparities in police officers' treatment of citizens through footage captured with body cameras worn by officers in the Oakland, California, Police Department. In the paper they subsequently published in the *Proceedings of the National Academy of Sciences,* the researchers reported that they had found police officers to be significantly less respectful and consistently ruder during routine traffic stops when the driver was black rather than white. Report coauthor Shelly Eberhardt clarified that "on the whole, officers were respectful to people," but "they were more respectful to whites than they were to blacks." Using automated scoring techniques, they rated more than 35,000 distinct utterances captured by body cameras worn by 245 officers during 981 stops in April 2014. For example, the statement, "All right my man. Do me a favor, just keep your hands on the wheel real quick" received a negative respect score of .51, while the statement, "Sorry to stop you. My name's Officer [name] with the Police Department" received a positive respect score of .84.

1. Are you convinced that these differences in treatment are the result of bias by the police? What else might explain them?
2. The researchers controlled (held constant) officer race, the severity of the driving violation, and other factors in order to isolate the effect of bias on the officers' behavior. Does you think this makes it more likely that the difference in respect reflected racial bias?

News source: Bromwich, Jonah Engel. 2017. "Police Are Less Respectful Toward Black Drivers, Report Finds." *The New York Times,* June 6, p. A12.

The solution is to know what the units of analysis and units of observation were in a study and to consider these in weighing the credibility of the researcher's conclusions. The goal is not to reject out of hand conclusions that refer to a level of analysis different from what was actually studied. Instead, the goal is to consider the likelihood that an ecological fallacy or a reductionist fallacy has been made when estimating the causal validity of the conclusions.

Cross-Sectional and Longitudinal Designs

Do you want to describe or understand social phenomena in the present? If you want to assess support for a candidate a month before a local election, you would need to collect your data at that time. If you want to describe the level of violence in Canadian communities, your focus might be on the current year. If the focus of your investigation is on the present or some other specific limited period, your research design will be cross-sectional. In cross-sectional research designs, all data are collected at one point in time. However, if you want to track changes in support for a candidate during the entire campaign period, or describe variation during the past decade in the level of violence in Canadian communities, you will need to collect data from the entire period you are investigating. In longitudinal research designs, data are collected at two or more points in time.

Therefore, the research question determines whether a researcher needs to collect cross-sectional or longitudinal data. If the research question concerns only the here and now, there is no need for longitudinal data. However, it is also important to recognize that any research question involving a causal analysis—about what causes what—creates an issue about change over time. Identifying the time order of effects—what happened first, and so on—is critical for developing a causal analysis but can be an insurmountable problem with a cross-sectional design. In longitudinal research designs, identification of the time order of effects can be quite straightforward.

Cross-Sectional Designs

Much of the research you have encountered so far in this text—the observations of computer use in Chapter 1, the surveys of binge drinking in Chapter 4 and of homeless persons in Chapter 5—has been cross-sectional. Although each of these studies took some time to carry out, researchers measured the actions, attitudes, and characteristics of respondents at only one point in time.

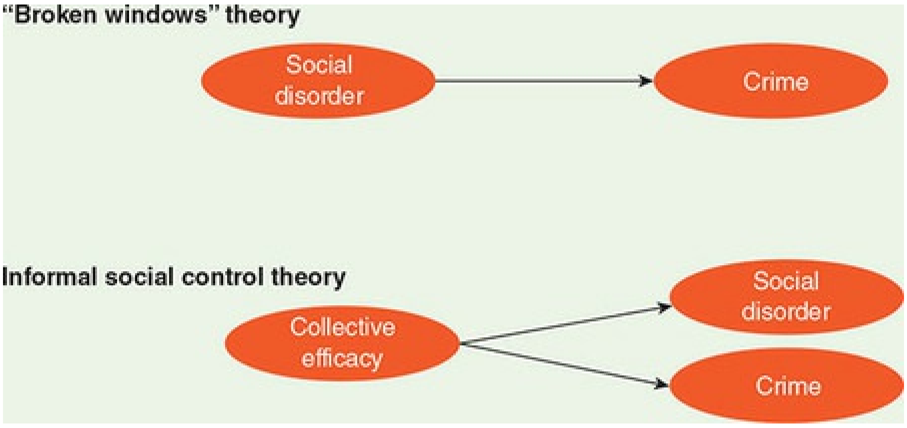
Sampson and Raudenbush (1999) used an ambitious cross-sectional design to investigate the effect of visible public social and physical disorder on the crime rate in Chicago neighborhoods. Their theoretical framework focused on the concept of informal social control—the ability of residents to regulate social activity in their neighborhoods through their collective efforts according to desired principles. The researchers believed that informal social control would vary between neighborhoods, and they hypothesized that the strength of informal social control, rather than just the visible signs of disorder, would explain variation in crime rates. Sampson and Raudenbush contrasted this prediction with the "broken windows" theory: the belief that signs of disorder themselves cause crime. In the theory Sampson and Raudenbush proposed, both visible disorder and crime were consequences of low levels of informal social control (measured with an index of collective efficacy). One did not cause the other (see Exhibit 6.3).

Cross-sectional research design: A study in which data are collected at only one point in time.

Longitudinal research design: A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.

Time order: A criterion for establishing a causal relation between two variables; the variation in the presumed cause (the independent variable) must occur before the variation in the presumed effect (the dependent variable).

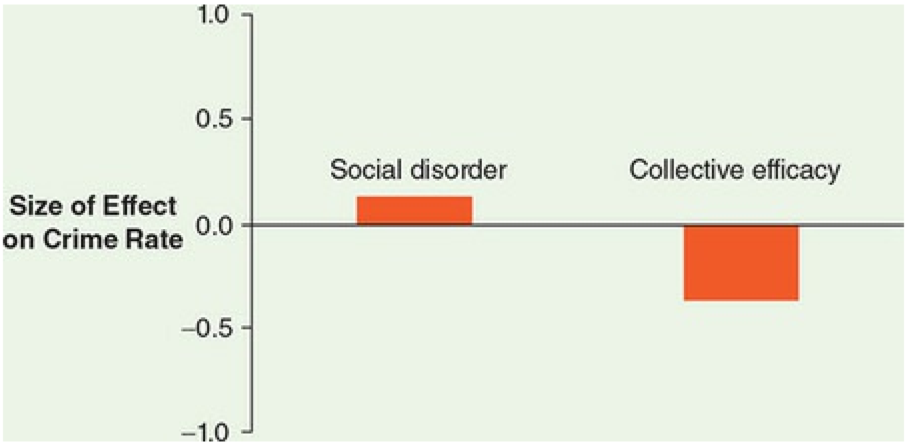
Exhibit 6.3 The Effect of Informal Social Control



Source: Based on Sampson and Raudenbush (1999).

Sampson and Raudenbush (1999) measured visible disorder through direct observation: Trained observers rode slowly around every street in 196 Chicago census tracts. Sampson and Raudenbush also conducted a survey of residents and examined police records. Both survey responses and police records were used to measure crime levels. The level of neighborhood informal social control and other variables were measured with the average resident responses to several survey questions. Both the crime rate and the level of social and physical disorder varied between neighborhoods in relation to the level of informal social control. Informal social control (collective efficacy) was a much more important factor in the neighborhood crime rate than was visible social and physical disorder, measured at the same time (see Exhibit 6.4).

Exhibit 6.4 Effect of Social Disorder and Collective Efficacy on Personal Violent Crime Rates



Source: Based on Sampson and Raudenbush (1999).

But note that we are left with a difficult question about the relations Sampson and Raudenbush identified between the variables they measured. Did neighborhoods that developed higher levels of informal social control then experience reductions in crime, or was it the other way around: Did neighborhoods that experienced a drop in crime then develop higher levels of informal social control? After all, if you are afraid to leave your apartment because you fear crime, you can't do a very good job of keeping an eye on things or attending community meetings. Maybe the crime reduction made residents feel safe to engage in more informal social control efforts, rather than just calling the police. Because of uncertainty like this, it is almost always better to use a longitudinal research design, rather than a cross-sectional research design, to answer questions about causal effects.

There are four special circumstances in which we can be more confident in drawing conclusions about time order—and hence conclusions about causality—on the basis of cross-sectional data. Because in these special circumstances the data can be ordered in time, they might even be thought of as longitudinal designs (Campbell 1992). These four special circumstances are as follows:

1. *The independent variable is fixed at some point before the variation in the dependent variable.* So-called demographic variables that are determined at birth—such as sex, race, and age—are fixed in this way. So are variables such as education and marital status, if we know when the value of cases on these variables was established and if we know that the value of cases on the dependent variable was set some time afterward. For example, say we hypothesize that education influences the type of job individuals have. If we know that respondents completed their education before taking their current jobs, we would satisfy the time order requirement even if we were to measure education at the same time we measure the type of job. However, if some respondents possibly went back to school as a benefit of their current jobs, the time

order requirement would not be satisfied.

2. *We believe that respondents can give us reliable reports of what happened to them or what they thought at some earlier point in time.* Julie Horney, D. Wayne Osgood, and Ineke Haen Marshall (1995) provide an interesting example of the use of such retrospective data. The researchers wanted to identify how criminal activity varies in response to changes in life circumstances. They interviewed 658 newly convicted male offenders sentenced to a Nebraska state prison. In a 45- to 90-minute interview, Horney et al. recorded each inmate's report of his life circumstances and of his criminal activities for the preceding 2 to 3 years. The researchers then found that criminal involvement was related strongly to adverse changes in life circumstances, such as marital separation or drug use. Retrospective data are often inadequate for measuring variation in past feelings, events, or behaviors, however, because we may have difficulty recalling what we have felt or what has happened in the past and what we do recall is likely to be influenced by what we feel in the present (Elliott, Holland, and Thomson 2008:229). For example, retrospective reports by both adult alcoholics and their parents appear to overestimate greatly the frequency of childhood problems (Vaillant 1995). People cannot report reliably the frequency and timing of many past events, from hospitalization to hours worked. However, retrospective data tend to be reliable when they concern major, persistent experiences in the past, such as what type of school someone went to or how a person's family was structured (Campbell 1992).
3. *Our measures are based on the records that contain information on cases in earlier periods.* Government, agency, and organizational records are excellent sources of time-ordered data after the fact. However, sloppy record keeping and changes in data collection policies can lead to inconsistencies, which must be considered. Another weakness of such archival data is that they usually contain measures of only a fraction of the variables that we think are important.
4. *We know that the value of the dependent variable was similar for all cases before the treatment.* For example, we may hypothesize that a training program (independent variable) improves the English-speaking abilities (dependent variable) of a group of recent immigrants. If we know that none of the immigrants could speak English before enrolling in the training program, we can be confident that any subsequent variation in their ability to speak English did not precede exposure to the training program. This is one way that traditional experiments establish time order: Two or more equivalent groups are formed before exposing one of them to some treatment.

Longitudinal Designs

In longitudinal research, data are collected that can be ordered in time. By measuring the value of cases on an independent variable and a dependent variable at different times, the researcher can determine whether variation in the independent variable precedes variation in the dependent variable.

In some longitudinal designs, the same sample (or panel) is followed over time; in other designs, sample members are rotated or completely replaced. The population from which the sample is selected may be defined broadly, as when a longitudinal survey of the general population is conducted. Or the population may be defined narrowly, as when the members of a specific age group are sampled at multiple points in time. The frequency of follow-up measurement can vary, ranging from a before-and-after design with just the one follow-up to studies in which various indicators are measured every month for many years.

Certainly, it is more difficult to collect data at two or more points in time than at one time. Quite frequently researchers simply cannot, or are unwilling to, delay completion of a study for even 1 year to collect follow-up data. But think of the many research questions that really should involve a much longer follow-up period: What is the impact of job training on subsequent employment? How effective is a school-based program in improving parenting skills? Under what conditions do traumatic experiences in childhood result in mental illness? It is safe to say that we will never have enough longitudinal data to answer many important research questions. Nonetheless, the value of longitudinal data is so great that every effort should be made to develop longitudinal research designs when they are appropriate for the research question asked. The following discussion of the three major types of longitudinal designs will give you a sense of the possibilities (see Exhibit 6.5).

Repeated cross-sectional designs (trend studies).

Studies that use a repeated cross-sectional design, also known as *trend studies,* have become fixtures of the political arena around election time. Particularly in presidential election years, we have all become accustomed to reading weekly, even daily, reports on the percentage of the population that supports each candidate. Similar polls are conducted to track sentiment on many other social issues. For example, a 1993 poll reported that 52% of adult Americans supported a ban on the possession of handguns, compared with 41% in a similar poll conducted in 1991. According to pollster Louis Harris, this increase indicated a “sea change” in public attitudes (cited in Barringer 1993). Another researcher said, “It shows that people are responding to their experience [of an increase in handgun-related killings]” (cited in Barringer 1993:A14).

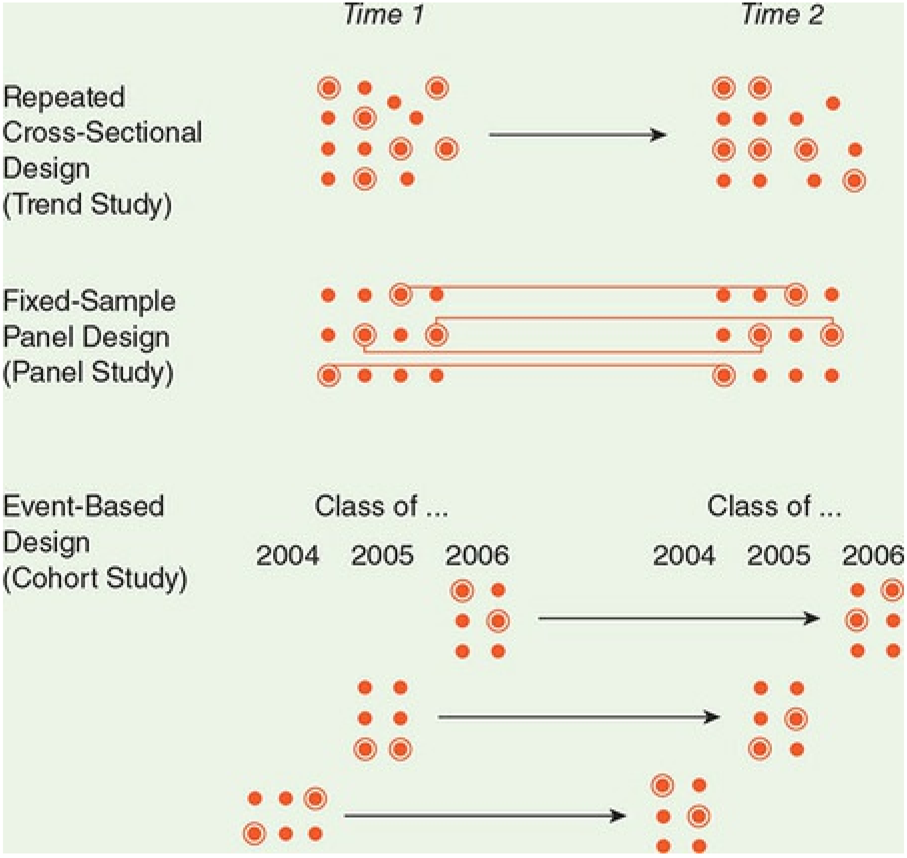
Repeated cross-sectional surveys are conducted as follows:

1. A sample is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some people leave the population and others enter it.
3. At Time 2, a different sample is drawn from this population.

Repeated cross-sectional design (trend study): A type of longitudinal study in which data are collected at two or more points in time from different samples of the same population.

These features make the repeated cross-sectional design appropriate when the goal is to determine whether a population has changed over time. Desmond, Papachristos, and Kirt (2016) used this design to study change over time in 911 calls. Has racial tolerance increased among Americans in the past 20 years? Are employers more likely to pay maternity benefits today than they were in the 1950s? These questions also concern the changes in the population as a whole, not just the changes in individuals within the population. We want to know whether racial tolerance increased in society, not whether this change was due to migration that brought more racially tolerant people into the country or to individual U.S. citizens becoming more tolerant. We are asking whether employers overall are more likely to pay maternity benefits today than they were yesterday, not whether any such increase was due to recalcitrant employers going out of business or to individual employers changing their maternity benefits. When we do need to know whether individuals in the population changed, we must turn to a panel design.

Exhibit 6.5 Three Types of Longitudinal Design


Time 1
Time 2
Repeated Cross-Sectional Design (Trend Study)
Fixed-Sample Panel Design (Panel Study)
Event-Based Design (Cohort Study)
Class of ...
2004
2005
2006
Class of ...
2004
2005
2006

Fixed-sample panel designs (panel studies).

Panel designs *allow* us to identify changes in individuals, groups, or whatever we are studying. This is the process for conducting fixed-sample panel designs:

1. A sample (called a panel) is drawn from a population at Time 1, and data are

collected from the sample.

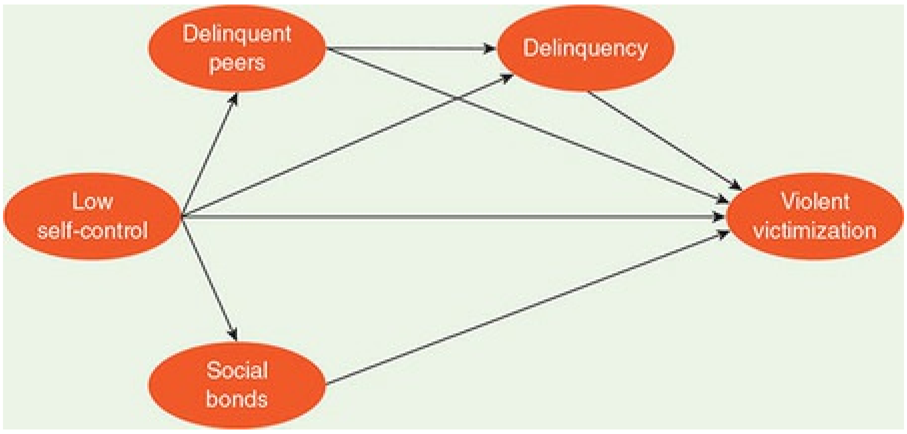
2. As time passes, some panel members become unavailable for follow-up, and the population changes.
3. At Time 2, data are collected from the same people as at Time 1 (the panel)—except for those people who cannot be located.

Because a panel design follows the same individuals, it is better than a repeated cross-sectional design for testing causal hypotheses. For example, Robert Sampson and John Laub (1990) used a fixed-sample panel design to investigate the effect of childhood deviance on adult crime. They studied a sample of white males in Boston when the subjects were between 10 and 17 years old and followed up when the subjects were in their adult years. Data were collected from multiple sources, including the subjects themselves and criminal justice records. Sampson and Laub (1990:614) found that children who had been committed to a correctional school for persistent delinquency were much more likely than were other children in the study to commit crimes as adults: 61% were arrested between the ages of 25 and 32, compared with 14% of those who had not been in correctional schools as juveniles. In this study, juvenile delinquency unquestionably occurred before adult criminality. If the researchers had used a cross-sectional design to study the past of adults, the juvenile delinquency measure might have been biased by memory lapses, by self-serving recollections about behavior as juveniles, or by loss of agency records.

Christopher Schreck, Eric Steward, and Bonnie Fisher (2006) wanted to identify predictors of adolescent victimization and wondered if the cross-sectional studies that had been conducted about victimization might have provided misleading results. Specifically, they suspected that adolescents with lower levels of self-control might be more prone to victimization and so needed to collect or find longitudinal data in which self-control was measured before experiences of victimization. The theoretical model they proposed to test included several other concepts that criminologists have identified as related to delinquency and that also might be influenced by levels of self-control: having delinquent peers, engaging in more delinquency, and being less attached to parents and school (see Exhibit 6.6). Schreck et al. analyzed data available from a panel study of delinquency and found that low self-control at an earlier time made it more likely that adolescents would subsequently experience victimization, even accounting for other influences. The researchers' use of a panel design allowed them to be more confident that the self-control–victimization relationship was causal than if they had used a cross-sectional design.

Fixed-sample panel design (panel study): A type of longitudinal study in which data are collected from the same individuals—the panel—at two or more points in time. In another type of panel design, panel members who leave are replaced with new members.

Exhibit 6.6 Schreck et al.'s (2006) Explanatory Model of Adolescent Victimization



Source: Copyright © 2006, Springer Science + Business Media, Inc. Reprinted with permission.

Despite their value in establishing time order of effects, panel studies are a challenge to implement successfully, so they often are not even attempted. There are two major difficulties:

1. *Expense and attrition.* It can be difficult, and very expensive, to keep track of individuals over a long period, and inevitably the proportion of panel members who can be located for follow-up will decline over time. Panel studies often lose more than one quarter of their members through attrition (Miller and Salkind 2002:321), and those who are lost were often not necessarily like those who remain in the panel. As a result, a high rate of subject attrition may mean that the follow-up sample will no longer be representative of the population from which it was drawn and may no longer provide a sound basis for estimating change. Subjects who were lost to follow-up may have been those who changed the most, or the least, over time. For example, between 5% and 66% of subjects are lost in substance abuse prevention studies, and the dropouts typically had begun the study with higher rates of tobacco and marijuana use (Snow, Tebes, and Arthur 1992:804).

   It does help to compare the baseline characteristics of those who are interviewed at follow-up with characteristics of those lost to follow-up. If these two groups of panel members were not very different at baseline, it is less likely that changes had anything to do with characteristics of the missing panel members. Even better, subject attrition can be reduced substantially if sufficient staff can be used to keep track of panel members. In their panel study, Sampson and Laub (1990) lost only 12% of the juveniles in the original sample (8% if you do not count those who had died).
2. *Subject fatigue.* Panel members may grow weary of repeated interviews and drop out of the study, or they may become so used to answering the standard questions in the survey that they start giving stock answers rather than actually thinking about their current feelings or actions (Campbell 1992). This problem is called subject fatigue.

> Fortunately, subjects do not often seem to become fatigued in this way, particularly if the research staff have maintained positive relations with the subjects. For example, at the end of an 18-month-long experimental study of housing alternatives for persons with mental illness who had been homeless, only 3 or 4 individuals (out of 93 who could still be located) refused to participate in the fourth and final round of interviews. The interviews took a total of about 5 hours to complete, and participants received about $50 for their time (Schutt, Goldfinger, and Penk 1997).

Subject fatigue: Problems caused by panel members growing weary of repeated interviews and dropping out of a study or becoming so used to answering the standard questions in the survey that they start giving stock or thoughtless answers.

Because panel studies are so useful, social researchers have developed increasingly effective techniques for keeping track of individuals and overcoming subject fatigue. But when resources do not permit use of these techniques to maintain an adequate panel, repeated cross-sectional designs usually can be employed at a cost that is not a great deal higher than that of a one-time-only cross-sectional study. The payoff in explanatory power should be well worth the cost.

Event-based designs (cohort studies).

In an event-based design, often called a *cohort study,* the follow-up samples (at one or more times) are selected from the same cohort—people who all have experienced a similar event or a common starting point. Examples include the following:

- *Birth cohorts*—those who share a common period of birth (those born in the 1940s, 1950s, 1960s, etc.)
- *Seniority cohorts*—those who have worked at the same place for about 5 years, about 10 years, and so on
- *School cohorts*—freshmen, sophomores, juniors, and seniors

An event-based design can be a type of repeated cross-sectional design or a type of panel design. In an event-based repeated cross-sectional design, separate samples are drawn from the same cohort at two or more different times. In an event-based panel design, the same individuals from the same cohort are studied at two or more different times. Comparing findings between different cohorts can help reveal the importance of the social or cultural context that the different cohorts experienced (Elliott et al. 2008:230).

Event-based research can improve identification of causal effects compared with cross-sectional designs. We can see this value of event-based research in a comparison between two studies that estimated the impact of public and private schooling on high school students' achievement test scores, only one of which used a cohort design. In a cross-sectional study, James Coleman, Thomas Hoffer, and Sally Kilgore (1982:68–69) compared standardized achievement test scores of high school sophomores and seniors in public, Catholic, and other private schools. The researchers found that test scores were higher in the private high schools (both Catholic and other) than in the public high schools. But was this difference a causal effect of private schooling? Perhaps the parents of higher-performing children were choosing to send them to private rather than to public schools. In other words, the higher achievement levels of private school students might have been in place before they started high school and might not have developed as a consequence of their high school education.

The researchers tried to reduce the impact of this problem by statistically controlling for a range of family background variables: family income, parents' education, race, number of siblings, number of rooms in the home, number of parents present, mother working, and other indicators of a family orientation to education. But some critics pointed out that even with all these controls for family background, the cross-sectional study did not ensure that the students had been comparable in achievement when they started high school.

Coleman and Hoffer (1987) thus went back to the high schools and studied the test scores of the former sophomores 2 years later, when they were seniors; in other words, the researchers used an event-based panel design (a cohort study). This time they found that the verbal and math achievement test scores of the Catholic school students had increased more over the 2 years than was the case for the public school students (it was not clear whether the scores of the other private school students had increased). Irrespective of students' initial achievement test scores, the Catholic schools seemed to "do more" for their students than did the public schools. This finding continued to be true even when the dropouts were studied. The researchers' causal conclusion rested on much stronger ground because they used a cohort study design.

Event-based design (cohort study): A type of longitudinal study in which data are collected at two or more points in time from individuals in a cohort.

Cohort: Individuals or groups with a common starting point. Examples include college class of 1997, people who graduated from high school in the 1980s, General Motors employees who started work between the years 1990 and 2000, and people who were born in the late 1940s or the 1950s (the baby boom generation).

Quantitative or Qualitative Causal Explanations

A cause is an explanation for some characteristic, attitude, or behavior of groups, individuals, or other entities (such as families, organizations, or cities) or for events. Most social scientists seek causal explanations that reflect tests of the types of hypotheses with which you are familiar (see Chapter 3): The independent variable is the presumed cause, and the dependent variable is the potential effect. For example, the study by Sampson and Raudenbush (2001) tested whether disorder in urban neighborhoods (the independent variable) leads to crime (the dependent variable). (As you know, they concluded that it didn't, at least not directly.) This type of causal explanation is termed *nomothetic.*

A different type of cause is the focus of some qualitative research (Chapter 10), some historical and comparative research (Chapter 15), and our everyday conversations about causes. In this type of causal explanation, termed *idiographic,* individual events or the behaviors of individuals are explained with a series of related, prior events. For example, you might explain a particular crime as resulting from several incidents in the life of the perpetrator that resulted in a tendency toward violence, coupled with stress resulting from a failed marriage and a chance meeting.

Quantitative (Nomothetic) Causal Explanations

A nomothetic causal explanation is one involving the belief that variation in an independent variable will be followed by variation in the dependent variable, when all other things are equal (ceteris paribus). Researchers who claim a causal effect from a nomothetic perspective have concluded that the value of cases on the dependent variable differs from what their value would have been in the absence of variation in the independent variable. For instance, researchers might claim that the likelihood of committing violent crimes is higher for individuals who were abused as children than it would be if these same individuals had not been abused as children. Or researchers might claim that the likelihood of committing violent crimes is higher for individuals exposed to media violence than it would be if these same individuals had not been exposed to media violence. The situation as it would have been in the absence of variation in the independent variable is termed the counterfactual (see Exhibit 6.7).

Of course, the fundamental difficulty with this perspective is that we never really know what would have happened at the same time to the same people (or groups, cities, etc.) if the independent variable had not varied—because it did (Shrout 2011:4–5). We can't rerun real-life scenarios (King, Keohane, and Verba 1994). We could observe the aggressiveness of people's behavior before and after they were exposed to media violence. But this comparison involves an earlier time, when, by definition, the people and their circumstances were not exactly the same.

But we do not need to give up hope! Far from it. We can design research to create conditions that are comparable indeed, so that we can confidently assert our conclusions *ceteris paribus*—other things being equal. We can examine the impact on the dependent variable of variation in the independent variable alone, even though we will not be able to compare the same people at the same time in exactly the same circumstances, except for the variation in the independent variable. And by knowing the ideal standard of comparability (the counterfactual), we can improve our research designs and strengthen our causal conclusions even when we cannot come so close to living up to the meaning of *ceteris paribus.*

Nomothetic causal explanation: An explanation that identifies common influences on a number of cases or events.

Ceteris paribus: Latin phrase meaning "other things being equal."

Causal effect (nomothetic perspective): When variation in one phenomenon, an independent variable, leads to or results, on average, in variation in another phenomenon, the dependent variable.

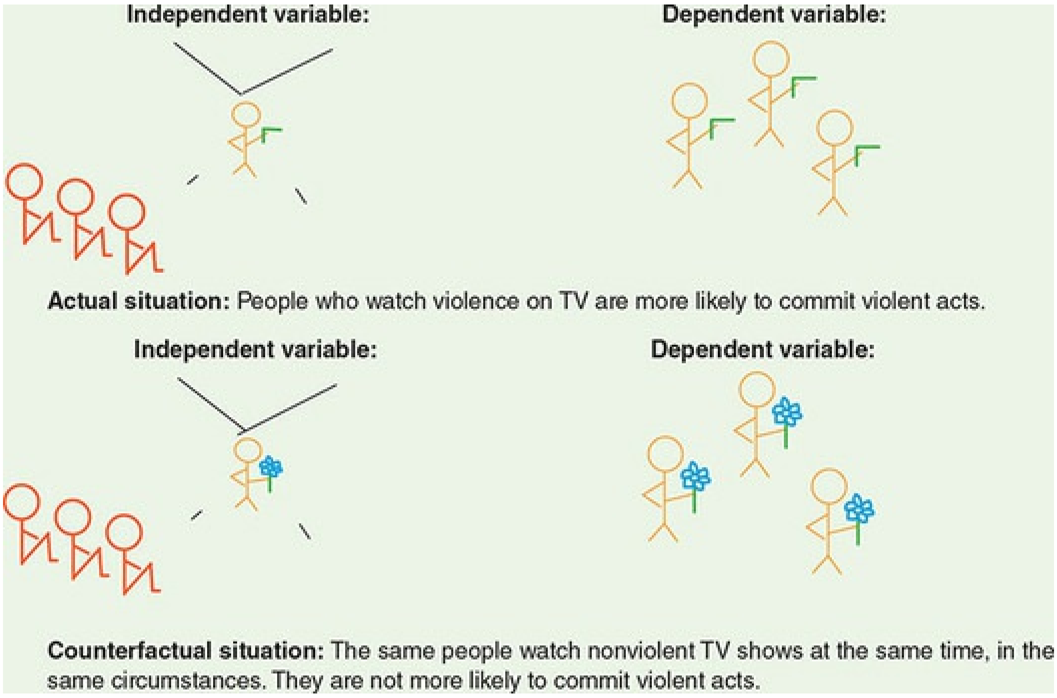
Example of a nomothetic causal effect: Individuals arrested for domestic assault tend to commit fewer subsequent assaults than do similar individuals who are accused in the same circumstances but not arrested.

Counterfactual: The situation that would have occurred if the subjects who were exposed to the treatment were not exposed but otherwise had had identical experiences to those they underwent during the experiment.

Idiographic causal explanation: An explanation that identifies the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event; may be termed an *individualist* or *historicist* explanation.

Quantitative researchers seek to test nomothetic causal explanations with either experimental or nonexperimental research designs. However, the way in which experimental and nonexperimental designs attempt to identify causes differs quite a bit. It is very hard to meet some of the criteria for achieving valid nomothetic causal explanations using a nonexperimental design. Most of the rest of this chapter is devoted to a review of these causal criteria and a discussion of how experimental and nonexperimental designs can help establish them.

Exhibit 6.7 The Counterfactual in Causal Research


Independent variable:
Dependent variable:
Actual situation: People who watch violence on TV are more likely to commit violent acts.
Independent variable:
Dependent variable:
Counterfactual situation: The same people watch nonviolent TV shows at the same time, in the same circumstances. They are not more likely to commit violent acts.

Qualitative (Idiographic) Causal Explanations

The other meaning of the term *cause* is one that we have in mind very often in everyday speech. This is idiographic causal explanation: the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event (Hage and Meeker 1988). An idiographic explanation also may be termed an *individualist* or a *historicist* explanation.

A causal effect from an idiographic perspective includes statements of initial conditions and then relates a series of events at different times that led to the outcome, or causal effect. This narrative or story is the critical element in an idiographic explanation, which may therefore be classified as narrative reasoning (Richardson 1995:200–201). Idiographic explanations focus on particular social actors, in particular social places, at particular social times (Abbott 1992). Idiographic explanations are also typically very concerned with context—with understanding the particular outcome as part of a larger set of interrelated circumstances. Idiographic explanations can thus be termed *holistic.*

> Causal effect (idiographic perspective): When a series of concrete events, thoughts, or actions results in a particular event or individual outcome.
>
> *Example of an idiographic causal effect:* An individual is neglected by her parents but has a supportive grandparent. She comes to distrust others, has trouble in school, is unable to keep a job, and eventually becomes homeless. She subsequently develops a supportive relationship with a shelter case manager, who helps her find a job and regain her housing (based on Hirsch 1989).

Elijah Anderson’s (1990) field research in a poor urban community produced a narrative account of how drug addiction can result in a downward slide into residential instability and crime:

> When addicts deplete their resources, they may go to those closest to them, drawing them into their schemes. . . . The family may put up with the person for a while. They provide money if they can. . . . They come to realize that the person is on drugs. . . . Slowly the reality sets in more and more completely, and the family becomes drained of both financial and emotional resources. . . . Close relatives lose faith and begin to see the person as untrustworthy and weak. Eventually the addict begins to “mess up” in a variety of ways, taking furniture from the house [and] anything of value. . . . Relatives and friends begin to see the person . . . as “out there” in the streets. . . . One deviant act leads to another. (pp. 86–87)

An idiographic explanation can also be developed from the narration of an individual. For example, Carole Cain interviewed AA (Alcoholics Anonymous) members about their experiences to learn how they construct their identities as alcoholics. In one interview, excerpted by Catherine Kohler Riessman (2008:71), “Hank” describes some of his experiences with drinking:

> One morning he found he could not get up even after several drinks. . . . When he did get up, he found AA, although he cannot remember how he knew where to go. . . . From the morning when he contacted AA, he did not drink again for over five years. . . . Life improved, he got himself in better shape and got back together with his wife. After several years, the marriage broke up again, and in anger with his wife, he went back to drinking for another five years.

An idiographic explanation such as Anderson’s or Cain’s pays close attention to time order and causal mechanisms. Nonetheless, it is difficult to make a convincing case that one particular causal narrative should be chosen over an alternative narrative (Abbott 1992). Does low self-esteem result in vulnerability to the appeals of drug dealers, or does a chance drug encounter precipitate a slide in self-esteem? Did drinking lead Hank to go to AA, or did he start drinking more because he knew it would push him to go to AA? Did his drinking start again because his marriage broke up, or did his orientation lead his wife to leave and to his renewed drinking? The prudent causal analyst remains open to alternative explanations.

Careers and Research

Jennifer A. Herbert, Crime Intelligence Analyst

Jennifer Herbert graduated with a double major in political science and justice studies from James Madison University in 2007. She had aspirations of becoming a police officer and eventually a detective. She was hired as a police officer after graduation, but she realized while at the police academy that she wanted to pursue the crime analysis career path in law enforcement. She became a crime analyst with a county police department. While working full time as an analyst, Herbert pursued an MA degree in intelligence at the American Military University. She then accepted a promotion to crime intelligence analyst with a county police division. After working as a crime analyst for six years, she cannot imagine doing anything else.

Every day is different working as a crime intelligence analyst. Some days, Herbert analyzes phone records and maps a suspect's whereabouts during the time of a crime. Other days, she maps the latest residential burglary trends and predicts where the next burglary will occur. She also completes research projects that examine quality-of-life issues for the community, including estimating crimes per 1,000 residents by neighborhood. Herbert's role as a crime analyst is equally important in preventing crime and in apprehension of offenders by patrol officers. She thinks the most rewarding part of her job is helping people who have been victimized by apprehending offenders and improving the quality of life for county residents. Herbert has some good advice for students interested in careers involving analysis:

> "If crime analysis interests you, ask your local police department if you can do an internship (paid or unpaid) to gain experience. Be sure to network with other crime analysts and let them know you are interested in pursuing a career in crime analysis. Courses in all forms of data analysis and geographic information systems (GIS) are almost essential to a career in crime analysis."

Idiographic explanation is deterministic, focusing on what caused a particular event to occur or what caused a particular case to change. As in nomothetic explanations, idiographic causal explanations can involve counterfactuals, by trying to identify what would have happened if a different circumstance had occurred. But unlike in nomothetic explanations, in idiographic explanations, the notion of a probabilistic relationship, an average effect, does not really apply. A deterministic cause has an effect in the case under consideration.

Criteria and Cautions For Nomothetic Causal Explanations

How the research is designed influences our ability to draw causal conclusions. In this section, I introduce the features that need to be considered in a research design to evaluate how well it can support nomothetic causal conclusions.

Three criteria must be considered when deciding whether a causal connection exists. When a research design leaves one or more of the criteria unmet, we may have some important doubts about causal assertions that the researcher may have made. These three criteria are generally considered the most important bases for identifying a nomothetic causal effect: (1) empirical association, (2) appropriate time order, and (3) nonspuriousness (Hammersley 2008:43). The features of experimental research designs are particularly well suited to meeting these criteria and for testing nomothetic causal explanations. However, we must also consider the degree to which these criteria are met when evaluating nonexperimental research that is designed to test causal hypotheses.

Two other issues that I introduce as “cautions” in this chapter are a bit different. They are not necessary for establishing that a causal connection exists, but they help us understand it better. If we have not identified a causal mechanism, the first caution, we do not understand fully *why* a causal connection exists. The second caution is to specify the context in which a causal effect occurs because by understanding *when* or *under what conditions* the causal effect occurs, we will understand better what that effect *is*. Providing information about both mechanism and context can considerably strengthen causal explanations (Hammersley 2008:44–45).

In the following subsections, I will indicate how researchers attempt to meet the three criteria and address the two cautions with both experimental and nonexperimental designs. Illustrations of experimental design features will use a 2002 study by M. Lyn Exum on the effect of intoxication and anger on aggressive intentions. Most illustrations of nonexperimental design features will be based on the study by Sampson and Raudenbush (1999) of neighborhood social control, which I introduced earlier.

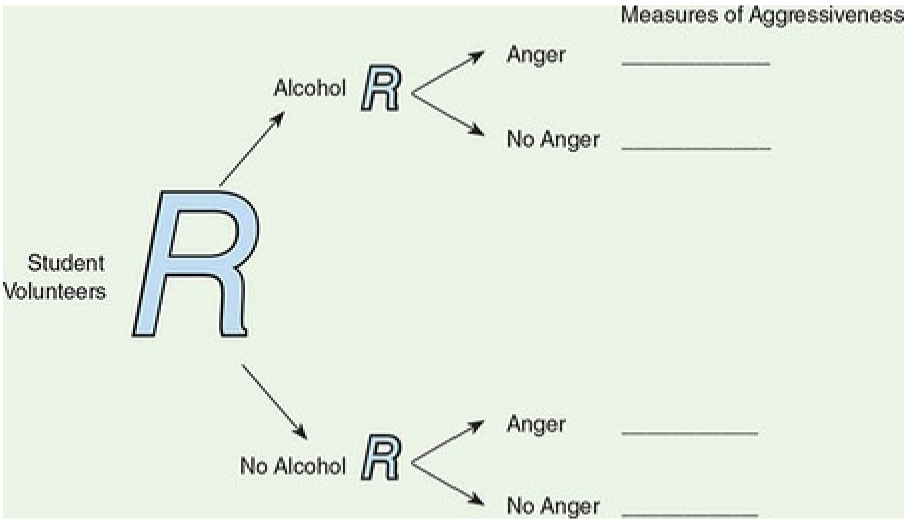
Exum (2002) and her assistants recruited 84 male students of legal drinking age at a mid-Atlantic university, using classroom announcements and fliers (women were not included because it was not possible for them to be screened for pregnancy before participation, as required by federal guidelines). Students who were interested in participating were given some background information that included the explanation that the study was about alcohol and cognitive skills. All participants were scheduled for individual appointments. When they arrived for the experiment, they completed a mood questionnaire and engaged in a meaningless video game. Those who were randomly assigned to the Alcohol condition were then given 1.5 ounces of 50% ethanol (vodka) per 40 pounds of body weight in orange juice (those in the No Alcohol condition just drank orange juice).

The other part of the experimental manipulation involved inducing anger among a randomly selected half of the participants. This was accomplished by the experimenter, who falsely accused the selected students of having come to the experiment 30 minutes late, informing them that as a consequence of their tardiness, they would not be paid as much for their time, and then loudly saying "bullshit" when the students protested. After these manipulations, the students read a fictional scenario involving the student and another man in a conflict about a girlfriend. The students were asked to rate how likely they would be to physically assault the other man, and what percentage of other male students they believed would do so (see Exhibit 6.8).

The students in the Alcohol condition (who were intoxicated) and had also been angered predicted that more students would react with physical aggression to the events depicted in the scenario (see Exhibit 6.9). The students in the four experimental conditions did not differ in their reports of their own likely aggressiveness, but Exum suggested this could be a result of the well-established phenomenon of *self-enhancement bias*—the tendency to evaluate oneself more positively than others. She concluded that she found mixed support for her hypothesis that alcohol increases violent decision making among persons who are angry.

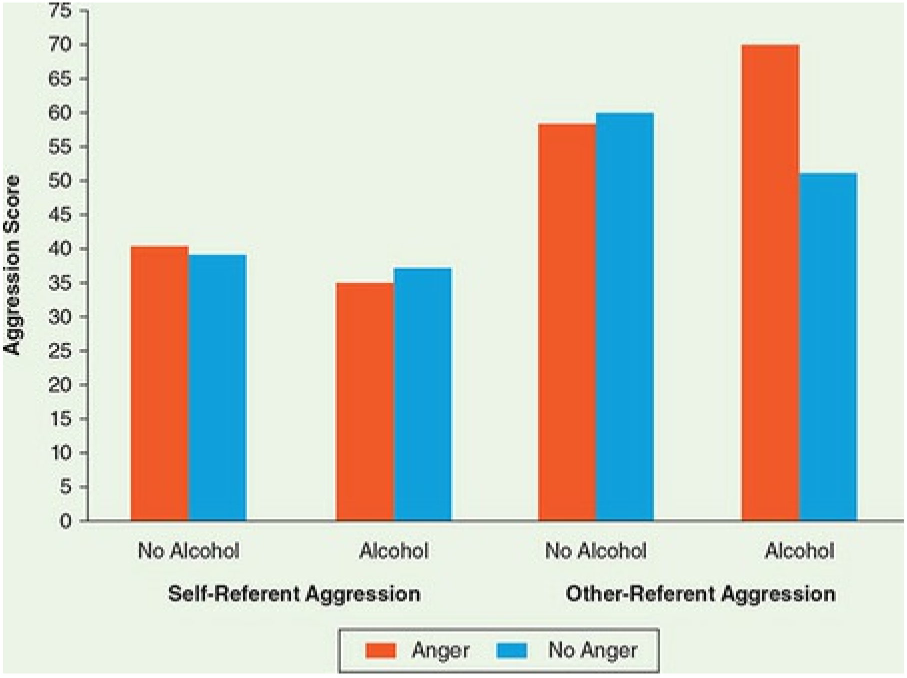
Was this causal conclusion justified? How confident can we be in its internal validity? Do you think that college students' reactions in a controlled setting with a fixed amount of alcohol are likely to be generalized to other people and settings? Does it help to know that Exum carefully confirmed that the students in the Alcohol condition were indeed intoxicated and that those in the Anger condition were indeed angry when they read the scenario? What about the causal conclusion by Sampson and Raudenbush (1999) that social and physical disorder does not directly cause neighborhood crime? Were you convinced? In the next sections, I will show how well the features of the research designs used by Exum and by Sampson and Raudenbush meet the criteria for nomothetic causal explanation, and thus determine the confidence we can place in their causal conclusions. I will also identify those features of a *true experiment* that make this research design particularly well suited to testing nomothetic causal hypotheses.

Exhibit 6.8 Experiment to Test the Effect of Intoxication and Anger on Intention to Aggress


Measures of Aggressiveness
Student Volunteers
R
Alcohol
R
Anger
No Anger
No Alcohol
R
Anger
No Anger

Source: Adapted from Exum, M. Lyn. 2002. "The Application and Robustness of the Rational Choice Perspective in the Study of Intoxicated and Angry Intentions to Aggress." Criminology 40:933–966. Reprinted with permission from the American Society of Criminology.

Exhibit 6.9 Measures of Aggression by Experimental Condition



Source: Adapted from Exum, M. Lyn. 2002. "The Application and Robustness of the Rational Choice Perspective in the Study of Intoxicated and Angry Intentions to Aggress." Criminology 40:933–966. Reprinted with permission from the American Society of Criminology.

Association

We say that there was an association between aggressive intentions and intoxication (for angry students) in Exum's (2002) experiment because the level of aggressive intentions varied according to whether students were intoxicated. An empirical (or observed) association between the independent and dependent variables is the first criterion for identifying a nomothetic causal effect.

We can determine whether an association exists between the independent and dependent variables in a true experiment because there are two or more groups that differ in their value on the independent variable. One group receives some "treatment"—such as reading a cathartic message—that manipulates the value of the independent variable. This group is termed the *experimental group.* In a simple experiment, there may be one other group that does not receive the treatment; it is termed the *control group.* The Exum study compared four groups created with two independent variables; other experiments may compare only two groups that differ in one independent variable, or more groups that represent multiple values of the independent variable or combinations of the values of more than two independent variables.

Association: A criterion for establishing a nomothetic causal relationship between two variables: Variation in one variable is related to variation in another variable.

In nonexperimental research, the test for an association between the independent and dependent variables is like that used in experimental research—seeing whether values of cases that differ on the independent variable tend to differ in the dependent variable. The difference with nonexperimental research designs is that the independent variable is not a treatment to which the researcher assigns some individuals. In their nonexperimental study of neighborhood crime, Sampson and Raudenbush (1999) studied the association between the independent variable (level of social and physical disorder) and the crime rate, but they did not assign individuals to live in neighborhoods with low or high levels of disorder.

But this is the time to be forewarned so that you can be forearmed: Have you heard the adage "correlation does not prove causation"? It highlights the fact that an association between two variables can occur for many reasons other than a causal effect of one on the other, including simple coincidence. Don't fall into that easy trap. The home team was winning more games and housing prices were rising? *Correlation does not prove causation!*

Time Order

Association is a necessary criterion for establishing a causal effect, but it is not sufficient. We must also ensure that the variation in the dependent variable occurred after the variation in the independent variable. This is the criterion of time order. Our research design determines our ability to determine time order.

Experimental Designs

In a true experiment, the researcher determines the time order. Exum (2002) first had some students drink alcohol and some experience the anger-producing manipulation and then measured their level of aggressive intentions. If we find an association between intoxication or anger and aggressiveness outside of an experimental situation, the criterion of time order may not be met. People who are more inclined to interpersonal aggression may be more likely than others to drink to the point of intoxication or to be angered by others in the first place. This would result in an association between intoxication and aggressive intentions, but the association would reflect the influence of being an aggressive person on drinking behavior rather than the other way around.

Nonexperimental Designs

You have already learned that nonexperimental research designs can be either cross-sectional or longitudinal. Because cross-sectional designs do not establish the time order of effects, their conclusions about causation must be more tentative. For example, although Sampson and Raudenbush (1999) found that lower rates of crime were associated with more informal social control (collective efficacy), their cross-sectional design could not establish directly that the variation in the crime rate occurred after variation in informal social control. Maybe it was a high crime rate that led residents to stop trying to exert much control over deviant activities in the neighborhood. It is difficult to discount such a possibility when only cross-sectional data are available, even though we can diagram hypothetical relations between the variables as if they are ordered in time (see Exhibit 6.10, panel 1).

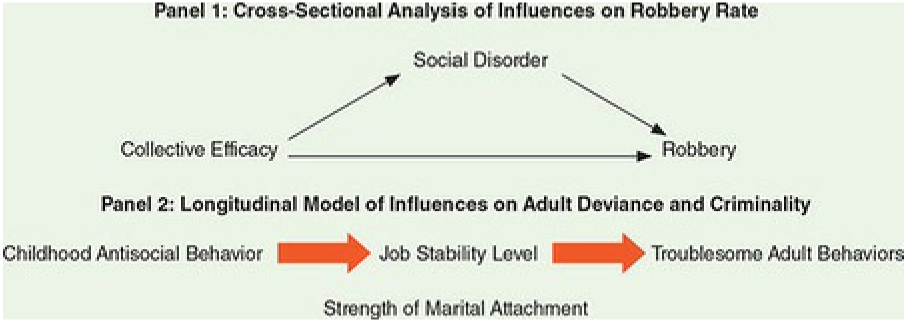
In contrast, Desmond, Papachristos, and Kirk's (2016) longitudinal study of the effects of police violence on reports of crime provided strong evidence of appropriate time order. Data on 911 calls were collected before widely publicized incidents of police violence as well as after them, so there's no question that the publicized incidents occurred before the change in 911 calls (see Exhibit 6.10, panel 2).

Nonspuriousness

Nonspuriousness is another essential criterion for establishing the existence of a causal effect of an independent variable on a dependent variable; in some respects, it is the most important criterion. We say that a relationship between two variables is not spurious when it is not caused by variation in a third variable. Always consider that an association between two variables could have been caused by something other than the presumed independent variable in that correlation—that is, it might be a spurious relationship rather than a causal one (see Exhibit 6.11). If we measure children's shoe sizes and their academic knowledge, for example, we will find a positive association. However, the association results from the fact that older children have larger feet as well as more academic knowledge. Shoe size does not cause knowledge, or vice versa.

Nonspuriousness: A criterion for establishing a causal relation between two variables; when a relationship between two variables is not caused by variation in a third variable.

Exhibit 6.10 Time Order in Nonexperimental Designs

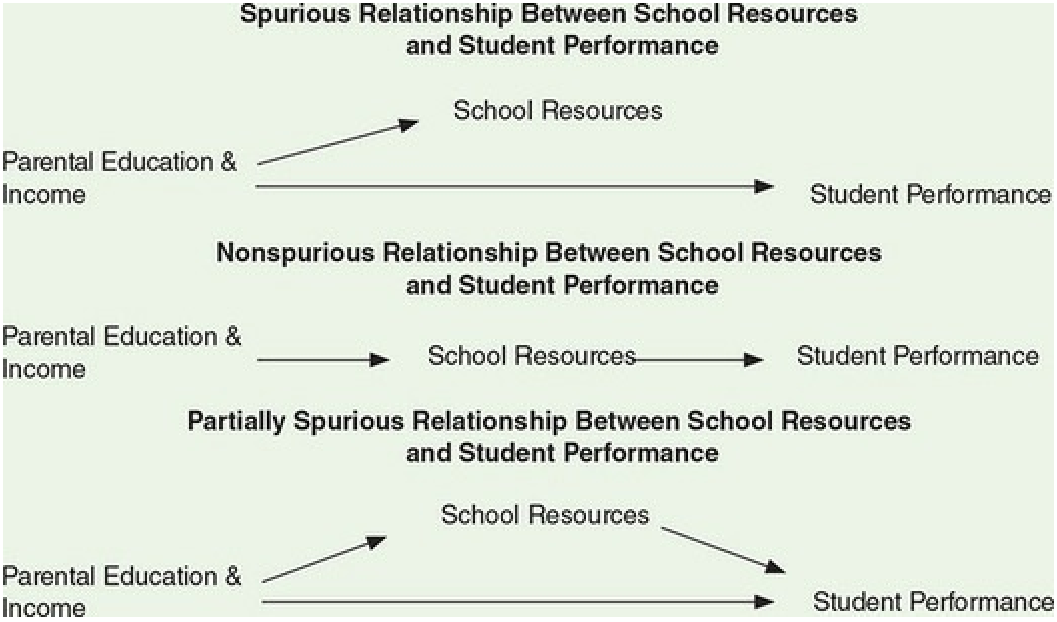

Panel 1: Cross-Sectional Analysis of Influences on Robbery Rate
Social Disorder
Collective Efficacy
Robbery
Panel 2: Longitudinal Model of Influences on Adult Deviance and Criminality
Childhood Antisocial Behavior
Job Stability Level
Troublesome Adult Behaviors
Strength of Marital Attachment

Sources: Based on Sampson and Raudenbush (1999); Desmond, Papachristos, and Kirk (2016).

Spurious relationship: A relationship between two variables that is caused by variation in a third variable.

Do storks bring babies? If you believe that correlation proves causation, then you might think so. The more storks that appear in certain districts in Holland, the more babies are born. But the association in Holland between number of storks and number of babies is spurious. In fact, both the number of storks and the birthrate are higher in rural districts than in urban districts. The rural or urban character of the districts (the extraneous variable) causes variation in the other two variables.

Exhibit 6.11 Spurious, Nonspurious, and Partially Spurious Relationships


Spurious Relationship Between School Resources and Student Performance
School Resources
Parental Education & Income
Student Performance
Nonspurious Relationship Between School Resources and Student Performance
Parental Education & Income
School Resources
Student Performance
Partially Spurious Relationship Between School Resources and Student Performance
School Resources
Parental Education & Income
Student Performance

If you think this point is obvious, consider our "Research That Matters" study about police violence and citizen crime reporting. Did the presumed independent variable (police violence) cause a change in the dependent variable (the likelihood of citizen crime reporting)? Desmond, Papachristos, and Kirk (2016) realized that variation in 911 calls between neighborhoods and over time could have been influenced by changes in the level of crime, so they took account of that possible source of spuriousness. Had you thought of that?

Randomization

A true experiment like Brad Bushman, Roy Baumeister, and Angela Stack's (1999) study of catharsis uses a technique called randomization to reduce the risk of spuriousness. Students in Bushman's experiment were asked to select a message to read by drawing a random number out of a bag. That is, the students were assigned randomly to a treatment condition. If students were assigned to only two groups, a coin toss could have been used (see Exhibit 6.12). Random assignment ensures that neither the students' aggressiveness nor any of their other characteristics or attitudes could influence which of the messages they read. As a result, the different groups are likely to be equivalent in all respects at the outset of the experiment. The greater the number of cases assigned randomly to the groups, the more likely that the groups will be equivalent in all respects. Whatever the preexisting sources of variation among the students, these could not explain why the group that read the procatharsis message became more aggressive, whereas the others didn't.

Statistical Control

A nonexperimental study such as Sampson and Raudenbush's (1999) cannot use random assignment to comparison groups to minimize the risk of spurious effects. Even if we wanted to, we couldn't randomly assign people to live in neighborhoods with different levels of informal social control. Instead, nonexperimental researchers commonly use an alternative approach to try to achieve the criterion of nonspuriousness. The technique of statistical control allows researchers to determine whether the relationship between the independent and dependent variables still occurs while we hold constant the values of other variables. If it does, the relationship could not be caused by variation in these other variables.

Extraneous variable: A variable that influences both the independent and dependent variables, creating a spurious association between them that disappears when the extraneous variable is controlled.

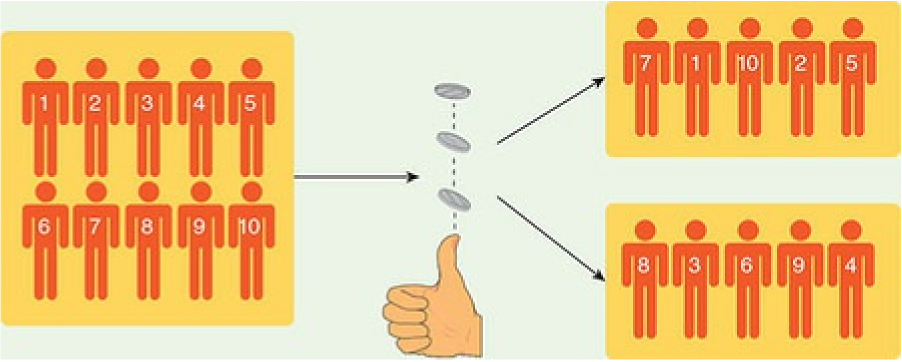
Randomization: The random assignment of cases, as by the toss of a coin.

Random assignment: A procedure by which each experimental subject is placed in a group randomly.

Statistical control: A method in which one variable is held constant so that the relationship between two (or more) other variables can be assessed without the influence of variation in the control variable.

Example of statistical control: In a different study, Sampson (1987) found a relationship between rates of family disruption and violent crime. He then classified cities by their level of joblessness (the control variable) and found that same relationship between the rates of family disruption and violent crime among cities with different levels of joblessness. Thus, the rate of joblessness could not have caused the association between family disruption and violent crime.

Exhibit 6.12 Random Assignment to One of Two Groups



Sampson and Raudenbush designed their study, in part, to determine whether the apparent effect of visible disorder on crime—the "broken windows" thesis—was spurious because of the effect of informal social control (see Exhibit 6.3). Exhibit 6.13 shows how statistical control was used to test this possibility. The data for all neighborhoods show that neighborhoods with much visible disorder had higher crime rates than did those with less visible disorder. However, when we examine the relationship between visible disorder and neighborhood crime rate separately for neighborhoods with high and low levels of informal social control (i.e., when we statistically control for social control level), we see that the crime rate no longer varies with visible disorder. Therefore, we must conclude that the apparent effect of broken windows was spurious because of the level of informal social control. Neighborhoods with low levels of social control were more likely to have high levels of visible social and physical disorder, and they were also more likely to have a high crime rate, but the visible disorder itself did not alter the crime rate.

We can strengthen our understanding of nomothetic causal connections, and increase the likelihood of drawing causally valid conclusions, by considering two cautions: the need to investigate causal mechanism and the need to consider the causal context. These two cautions are emphasized in the definition of idiographic causal explanation, with its attention to the sequence of events and the context in which they happen, but here I will limit my discussion to research oriented toward nomothetic causal explanations.

Mechanism

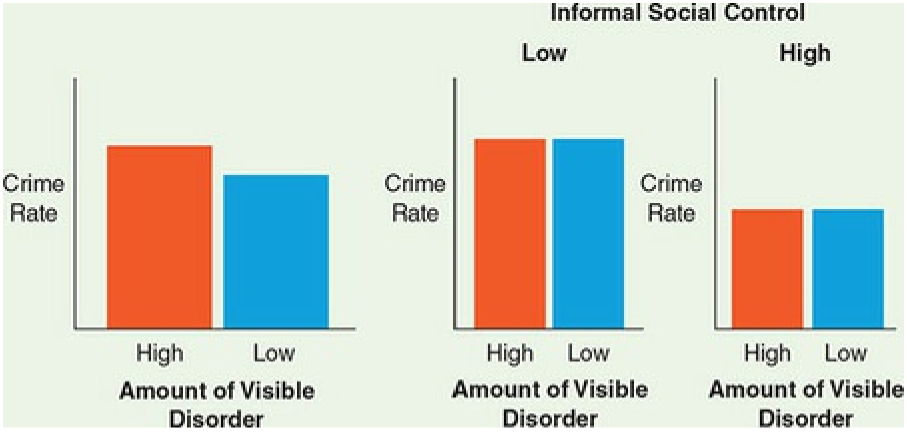
A causal mechanism is some process that creates the connection between variation in an independent variable and the variation in the dependent variable the independent variable is hypothesized to cause (Cook and Campbell 1979:35; Marini and Singer 1988). Many social scientists (and scientists in other fields) argue that no nomothetic causal explanation is adequate until a causal mechanism is identified (Costner 1989; Hedström and Swedberg 1998). In statistical analysis, variables that involve a mechanism are termed mediators.

Mechanism: A discernible process that creates a causal connection between two variables.

Mediator: A variable involved in a causal mechanism (intervening variable).

Our confidence in causal conclusions based on nonexperimental research increases with identification of a causal mechanism (Shrout 2011:15–16). Such mechanisms help us understand how variation in the independent variable results in variation in the dependent variable. For example, in a study that reanalyzed data from Sheldon Glueck and Eleanor Glueck's (1950) pathbreaking study of juvenile delinquency, Sampson and Laub (1994) found that children who grew up with structural disadvantages such as family poverty and geographic mobility were more likely to become juvenile delinquents. Why did this occur? Sampson and Laub's (1994) analysis indicated that these structural disadvantages led to lower levels of informal social control in the family (less parent–child attachment, less maternal supervision, and more erratic or harsh discipline). In turn, lower levels of informal social control resulted in a higher probability of delinquency (see Exhibit 6.14). Informal social control thus intervened in the relationship between structural disadvantage and juvenile delinquency.

Exhibit 6.13 The Use of Statistical Control to Reduce Spuriousness



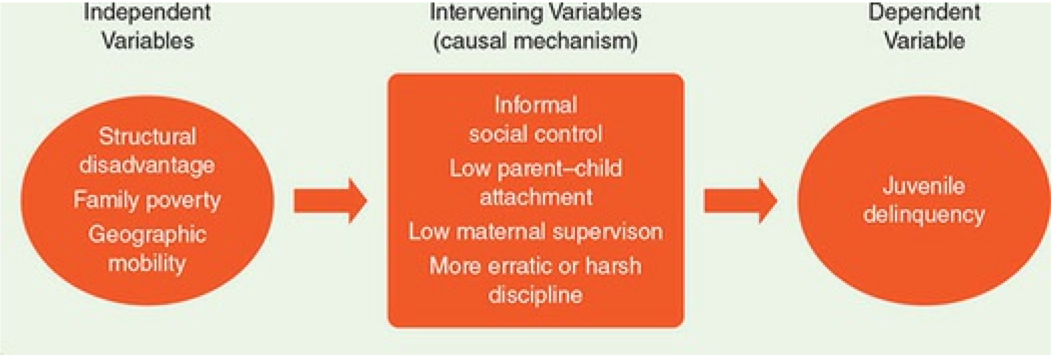
In their study of deterrence of spouse abuse (introduced in Chapter 2), Lawrence Sherman and Richard Berk (1984) designed follow-up experiments to test or control several causal mechanisms that they wondered about after their first experiment: Did recidivism decrease for those who were arrested for spouse abuse *because* of the exemplary work of the arresting officers? Did recidivism increase for arrestees *because* they experienced more stressors with their spouses as time passed? Investigating these and other possible causal mechanisms enriched Sherman and Berk's eventual explanation of how arrest influences recidivism.

Of course, you might ask why structural disadvantage tends to result in lower levels of family social control or how family social control influences delinquency. You could then conduct research to identify the mechanisms that link, for example, family social control and juvenile delinquency. (Perhaps the children feel they're not cared for, so they become less concerned with conforming to social expectations.) This process could go on and on. The point is that identification of a mechanism through which the independent variable influences the dependent variable increases our confidence in the conclusion that a causal connection does indeed exist. However, identification of a causal mechanism—one or more mediating variables—in turn requires concern for the same causal criteria that we consider when testing the original relationship, including time order and nonspuriousness (Shrout 2011:15–21).

Context

Do the causal processes in which we are interested vary across neighborhoods? Among organizations? Across regions? Over time? For different types of people? When relationships between variables differ across geographic units such as counties or across other social settings, researchers say there is a contextual effect. Identification of the context in which a causal relationship occurs can help us understand that relationship. The changes in the crime rate with which we began this chapter differed for blacks and whites, for youth and adults, and in urban and rural areas (Ousey and Lee 2004: 359–360). These contextual effects suggest that single-factor explanations about these changes are incorrect (Rosenfeld 2004:89). In statistical analysis, variables that identify contexts for the effects of other variables are termed moderators.

Exhibit 6.14 Intervening Variables in Nonexperimental Research: Structural Disadvantage and Juvenile Delinquency


Independent Variables
Intervening Variables (causal mechanism)
Dependent Variable
Structural disadvantage
Family poverty
Geographic mobility
Informal social control
Low parent–child attachment
Low maternal supervison
More erratic or harsh discipline
Juvenile delinquency

Source: Based on Sampson and Raudenbush (1999).

Sampson and Laub (1993) found support for a contextual effect in their study of 538,000 juvenile justice cases in 322 U.S. counties: In counties having a relatively large underclass and poverty concentrated among minorities, juvenile cases were more likely to be treated harshly. These relationships occurred for both African American and white juveniles, but they were particularly strong for African Americans. The results of this research suggest the importance of considering social context when examining criminal justice processes (see also Dannefer and Schutt 1982; Schutt and Dannefer 1988).

Exum (2002) tested the effect of intoxication itself on aggressive intentions, but also in relation to whether students were angry or not. She found that only in the context of being angry did alcohol lead the students to express more aggressive intentions (relative to how they said other students would act; see Exhibit 6.9).

Context was also important in Sherman and Berk's (1984) research on domestic violence.

Arrest was less effective in reducing subsequent domestic violence in cities with high levels of unemployment than in cities with low levels of unemployment. This seemed to be more evidence of the importance of individuals having a “stake in conformity” (Berk et al. 1992). Awareness of contextual differences helps us make sense of the discrepant findings from local studies. Always remember that the particular cause on which we focus in a given research design may be only one among a set of interrelated factors required for the effect; when we account for context, we specify these other factors (Hage and Meeker 1988; Papineau 1978).

Contextual effect: A variation in relationships of dependent with independent variables between geographic units or other social settings.

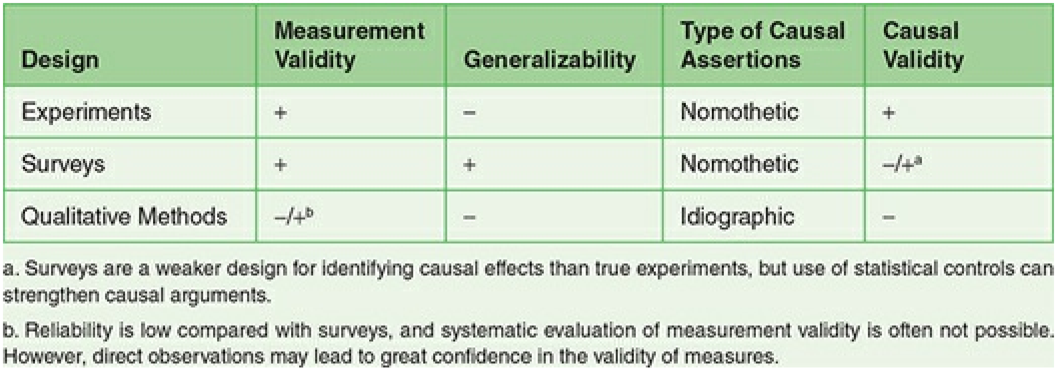
Context: A set of interrelated circumstances that alters a relationship between other variables or social processes.

Moderator: A variable that identifies a context for the effect of other variables.

Comparing Research Designs

The central features of the basic social research designs you will study in detail in the next chapters—experiments, surveys, and qualitative methods—provide distinct perspectives even when used to study the same social processes. Comparing subjects randomly assigned to a treatment and to a comparison group, asking standard questions of the members of a random sample, or observing while participating in a natural social setting involves markedly different decisions about measurement, causality, and generalizability. As you can see in Exhibit 6.15, not one of these research designs can reasonably be graded as superior to the others in all respects, and each varies in its suitability to different research questions and goals. Choosing among them for a particular investigation requires consideration of the research problem, opportunities and resources, prior research, philosophical commitments, and research goals.

Exhibit 6.15 Comparison of Research Methods



Design	Measurement Validity	Generalizability	Type of Causal Assertions	Causal Validity
Experiments	+	–	Nomothetic	+
Surveys	+	+	Nomothetic	–/+[a]
Qualitative Methods	–/+[b]	–	Idiographic	–

a. Surveys are a weaker design for identifying causal effects than true experiments, but use of statistical controls can strengthen causal arguments.

b. Reliability is low compared with surveys, and systematic evaluation of measurement validity is often not possible. However, direct observations may lead to great confidence in the validity of measures.

Experimental designs are strongest for testing nomothetic causal hypotheses and are most appropriate for studies of treatment effects. Research questions that are believed to involve basic social psychological processes are most appealing for laboratory studies because the problem of generalizability is reduced. Random assignment reduces the possibility of preexisting differences between treatment and comparison groups to small, specifiable, chance levels; therefore, many of the variables that might create a spurious association are controlled. But despite this clear advantage, an experimental design requires a degree of control that cannot always be achieved outside the laboratory. It can be difficult to ensure in real-world settings that a treatment is delivered as intended and that other influences do not intrude. As a result, what appears to be a treatment effect or noneffect may be something else altogether. Field experiments thus require careful monitoring of the treatment process. Unfortunately, most field experiments also require more access arrangements and financial resources than can often be obtained.

Laboratory experiments permit much more control over conditions, but at the cost of less generalizable findings. People must volunteer for most laboratory experiments, and so there is a good possibility that experimental subjects differ from those who do not volunteer.

Ethical and practical constraints limit the types of treatments that can be studied experimentally (you can't assign social class or race experimentally). The problem of generalizability in an experiment using volunteers lessens when the object of investigation is an orientation, behavior, or social process that is relatively invariant among people, but it is difficult to know which orientations, behaviors, or processes are so invariant. If a search of the research literature on the topic identifies many prior experimental studies, the results of these experiments will suggest the extent of variability in experimental effects and point to the unanswered questions about these effects.

Both surveys and experiments typically use standardized, quantitative measures of attitudes, behaviors, or social processes. Closed-ended questions are most common and are well suited for the reliable measurement of variables that have been studied in the past and whose meanings are well understood (see Chapter 4). Of course, surveys often include measures of many more variables than are included in an experiment (Chapter 8), but this feature is not inherent in either design. Phone surveys may be quite short, whereas some experiments can involve very lengthy sets of measures (see Chapter 7). The set of interview questions we used at baseline in the Boston housing study (Schutt 2011b) mentioned in Chapters 10 and 12, for example, required more than 10 hours to complete. The level of funding for a survey will often determine which type of survey is conducted and thus the length of the questionnaire.

Most social science surveys rely on random sampling for their selection of cases from some larger population, and this feature makes them preferable for descriptive research that seeks to develop generalizable findings (see Chapter 5). However, survey questionnaires can only measure what respondents are willing to report verbally; questionnaires may not be adequate for studying behaviors or attitudes that are regarded as socially unacceptable. Surveys are also often used to test hypothesized causal relationships. When variables that might create spurious relationships are included in the survey, they can be controlled statistically in the analysis and thus eliminated as rival causal influences.

Conclusions

Causation and the means for achieving causally valid conclusions in research is the last of the three legs on which the validity of research rests. In this chapter, you have learned about two alternative meanings of causation (nomothetic and idiographic). You have studied the three criteria and two cautions used to evaluate the extent to which particular research designs may achieve causally valid findings. You have learned how our ability to meet these criteria is shaped by research design features such as units of analysis, use of a cross-sectional or longitudinal design, and use of randomization or statistical control to deal with the problem of spuriousness. You have also seen why the distinction between experimental and nonexperimental designs has so many consequences for how, and how well, we are able to meet nomothetic criteria for causation.

I should reemphasize that the results of any particular study are part of an always-changing body of empirical knowledge about social reality. Thus, our understandings of causal relationships are always partial. Researchers always wonder whether they have omitted some relevant variables from their controls, whether their experimental results would differ if the experiment were conducted in another setting, or whether they have overlooked a critical historical event. But by using consistent definitions of terms and maintaining clear standards for establishing the validity of research results—and by expecting the same of others who do research—social researchers can contribute to a growing body of knowledge that can reliably guide social policy and social understanding.

When you read the results of a social scientific study, you should now be able to evaluate critically the validity of the study's findings. If you plan to engage in social research, you should now be able to plan an approach that will lead to valid findings. And with a good understanding of three dimensions of validity (measurement validity, generalizability, and causal validity) under your belt, and with sensitivity also to the goal of *authenticity,* you are ready to focus on the major methods of data collection used by social scientists.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Association 205
Causal effect (idiographic perspective) 201
Causal effect (nomothetic perspective) 200
Ceteris paribus 200
Cohort 199
Context 211
Contextual effect 211
Counterfactual 200
Cross-sectional research design 193
Ecological fallacy 190
Emergence 191
Event-based design (cohort study) 199
Extraneous variable 208
Fixed-sample panel design (panel study) 197
Idiographic causal explanation 201
Longitudinal research design 193
Mechanism 209
Mediator 209
Moderator 211
Nomothetic causal explanation 200
Nonspuriousness 206
Random assignment 208
Randomization 208
Reductionist fallacy (reductionism) 191
Repeated cross-sectional design (trend study) 196
Spurious relationship 207
Statistical control 208
Subject fatigue 198
Time order 193
Units of analysis 189
Units of observation 190

Highlights

- We do not fully understand the variables in a study until we know to which units of analysis—what level of social life—they refer.
- Invalid conclusions about causality may occur when relationships between variables measured at the group level are assumed to apply at the individual level (the ecological fallacy) and when relationships between variables measured at the level of individuals are assumed to apply at the group level (the reductionist fallacy). Nonetheless, many research questions point to relationships at

multiple levels and so may profitably be investigated at multiple units of analysis.

- Longitudinal designs are usually preferable to cross-sectional designs for establishing the time order of effects. Longitudinal designs vary in whether the same people are measured at different times, how the population of interest is defined, and how frequently follow-up measurements are taken. Fixed-sample panel designs provide the strongest test for the time order of effects, but they can be difficult to carry out successfully because of their expense as well as subject attrition and fatigue.
- Causation can be defined in either nomothetic or idiographic terms. Nomothetic causal explanations deal with effects on average. Idiographic causal explanations deal with the sequence of events that led to a particular outcome.
- The concept of nomothetic causal explanation relies on a comparison. The value of cases on the dependent variable is measured after they have been exposed to variation in an independent variable. This measurement is compared with what the value of cases on the dependent variable would have been if they had not been exposed to the variation in the independent variable (the counterfactual). The validity of nomothetic causal conclusions rests on how closely the comparison group comes to the ideal counterfactual.
- From a nomothetic perspective, three criteria are generally viewed as necessary for identifying a causal relationship: (1) association between the variables, (2) proper time order, and (3) nonspuriousness of the association. In addition, the basis for concluding that a causal relationship exists is strengthened by the identification of a causal mechanism and the context for the relationship.
- Association between two variables is in itself insufficient evidence of a causal relationship. This point is commonly made with the expression "correlation does not prove causation."
- Experiments use random assignment to make comparison groups as similar as possible at the outset of an experiment to reduce the risk of spurious effects resulting from extraneous variables.
- Nonexperimental designs use statistical controls to reduce the risk of spuriousness. A variable is controlled when it is held constant so that the association between the independent and dependent variables can be assessed without being influenced by the control variable.
- Ethical and practical constraints often preclude the use of experimental designs.
- Idiographic causal explanations can be difficult to identify because the starting and ending points of particular events and the determination of which events act as causes in particular sequences may be ambiguous.

Discussion Questions

1. There's a lot of "sound and fury" in the social science literature about units of analysis and levels of explanation. Some social researchers may call another a *reductionist* if the latter explains a problem such as substance abuse as caused by "lack of self-control." The idea is that the behavior requires consideration of social structure—a group level of analysis rather than an individual level of analysis. Another researcher may be said to commit an *ecological fallacy* if she assumes that group-level characteristics explain behavior at the individual level (such as saying that "immigrants are more likely to commit crime" because the *neighborhoods* with higher proportions of immigrants have higher crime rates). Do you favor causal explanations at the individual or the group (or social structural) level? If you were forced to mark on a scale from 0 to 100 the percentage of crime that results from problems with individuals rather than from problems with the settings in which they live, where would you make your mark? Explain your decision.
2. Researchers often try to figure out how people have changed over time by conducting a cross-sectional survey of people of different ages. The idea is that if people who are in their 60s tend to be happier than people who are in their 20s, it is because people tend to "become happier" as they age. But maybe people who are in their 60s now were just as happy when they were in their 20s, and people in their 20s now will be just as unhappy when they are in their 60s. (That's called a *cohort effect.*) We can't be sure unless we conduct a panel or cohort study (survey the same people at different ages). What, in your experience, are the major differences between the generations today in social attitudes and behaviors? Which would you attribute to changes as people age, and which to differences between cohorts in what they have experienced (such as common orientations among baby boomers)? Explain your reasoning.
3. The chapter begins with some alternative explanations for recent changes in the crime rate. Which of the explanations make the most sense to you? Why? How could you learn more about the effect on crime of one of the "causes" you have identified in a laboratory experiment? What type of study could you conduct in the community to assess its causal impact?
4. This chapter discusses both experimental and nonexperimental approaches to identifying causes. What are the advantages and disadvantages of both approaches for addressing each of the three criteria and two cautions identified for causal explanations?
5. Construct an idiographic causal explanation for a recent historical or personal event. For example, what was the sequence of events that led to the outcome of the 2016 U.S. presidential election? What was the sequence of events that led to the replacement of Travis Kalanick as the CEO of Uber? (I know, you thought this would be easy.)

Practice Exercises

1. The study site contains lessons on units of analysis and the related problems of ecological fallacy and reductionism in the interactive exercises. Choose the "Units of Analysis" lesson from the main menu. It describes several research projects and asks you to identify the units of analysis in each. Then it presents several conclusions for particular studies and asks you to determine whether an error has been made.
2. Thomas Rotolo and Charles Tittle (2006) were puzzled by a contradictory finding about the relationship between city population size and crime rates: The results of most cross-sectional studies differ from those typically obtained in longitudinal studies. To test different causal hypotheses about this relationship, they obtained data about 348 cities that had at least 25,000 residents and adequate data about crime at four time points from 1960 to 1990. Let's review different elements of their arguments and use them to review criteria for causality.
    a. Cross-sectional studies tend to find that cities with more people have a higher crime rate. What criterion for causality does this establish? Review each of the other criteria for causality and explain what noncausal bases they suggest could account for this relationship.
    b. Some observers have argued that larger cities have higher rates of crime because large size leads to less social integration, which in turn leads to more crime. Which causal criterion does this explanation involve? How much more confident would you be that there is a causal effect of size on crime if you knew that this other relationship occurred also? Explain your reasoning.
    c. Evidence from longitudinal studies has been more mixed, but most do not find a relationship between city size and the crime rate. What do you think could explain the lack of a longitudinal relationship despite the cross-sectional relationship? Explain.
    d. Some observers have proposed that the presence of transients in large cities is what leads to higher crime rates because transients (those who are not permanent residents) are more likely to commit crimes. What causal criterion does this involve? Draw a diagram that shows your reasoning.
    e. In their analysis, Rotolo and Tittle (2006) control for region because they suggest that in regions that are traditionally very urban, people may be accustomed to rapid patterns of change, whereas in newly urbanizing regions this may not be the case. What type of causal criterion would region be? What other factors like this do you think the analysis should consider? Explain your reasoning.
    f. Now you can examine the Rotolo and Tittle (2006) article online (if your library subscribes to the *Journal of Quantitative Criminology*) and read the details.
3. Search Sociological Abstracts or another index to the social science literature for several articles on studies using any type of longitudinal design. You will be searching for article titles that use words such as *longitudinal, panel, trend,* or *over time.* How successful were the researchers in carrying out the design? What steps did the researchers who used a panel design take to minimize panel attrition? How convinced are you by those using repeated cross-sectional designs that they have identified a process of change in individuals? Did any researchers use retrospective questions? How did they defend the validity of these measures?

Ethics Questions

1. Randomization is a key feature of experimental designs that are often used to investigate the efficacy of new treatments for serious and often incurable, terminal diseases. What ethical issues do these techniques raise in studies of experimental treatments for incurable, terminal diseases? Would you make an ethical argument that there are situations when it is *more* ethical to use random assignment than usual procedures for deciding whether patients receive a new treatment?
2. You learned in this chapter that Sampson and Raudenbush (1999) had observers drive down neighborhood streets in Chicago and record the level of disorder they observed. What should have been the observers' response if they observed a crime in progress? What if they just suspected that a crime was going to occur? What if the crime was a drug dealer interacting with a driver at the curb? What if it was a prostitute soliciting a customer? What, if any, ethical obligation does a researcher studying a neighborhood have to residents in that neighborhood? Should research results be shared at a neighborhood forum?
3. Exum's (2002) experimental manipulation included having some students drink to the point of intoxication. This was done in a carefully controlled setting, with a measured amount of alcohol, and the students who were intoxicated were kept in a room after the experiment was finished until they were sober. Exum also explained the experiment to all the students when they finished the experiment. If you were a student member of your university's Institutional Review Board, would you vote to approve a study with these features? Why or why not? Would you ban an experiment like this involving alcohol altogether, or would you set even more stringent criteria? If the latter, what would those criteria be? Do you think Exum should have been required to screen prospective female students for pregnancy so that some women could have been included in the study (who were not pregnant)? Can you think of any circumstances in which you would allow an experiment involving the administration of illegal drugs?

Web Exercises

1. Go to the Disaster Center website, www.disastercenter.com/crime. Review the crime rate nationally, and, by picking out links to state reports, compare the recent crime rates in two states. Report on the prevalence of the crimes you have examined. Propose a causal explanation for variation in crime between states over time, or both. What research design would you propose to test this explanation? Explain.
2. Go to the Crime Stoppers USA (CSUSA) website at www.crimestoppersusa.org. Check out "Profile" and then "FAQ." How is CSUSA "fighting crime"? What does CSUSA's approach assume about the cause of crime? Do you think CSUSA's approach to fighting crime is based on valid conclusions about causality? Explain.
3. What are the latest trends in crime? Write a short statement after inspecting the FBI's Uniform Crime Reports at www.fbi.gov (go to the "Services" section and then "UCR"; select "Crime Stats for 20__" and then pick specific categories to review).

Video Interview Questions

Listen to the researcher interview for Chapter 6 at edge.sagepub.com/schutt9e.

1. How does Professor Neale distinguish quantitative longitudinal research from qualitative longitudinal research? Which is more appealing to you? Why?
2. What unique advantages of longitudinal research does Professor Neale highlight? Which research questions are most suited to longitudinal research? Please explain your reasoning.

SPSS Exercises

We can use the GSS2016 data to learn how causal hypotheses can be evaluated with nonexperimental data.

1. Specify four hypotheses in which CAPPUN is the dependent variable and the independent variable is also measured with a question in the 2016 GSS. The independent variables should have no more than 10 valid values (check the variable list).
    a. Inspect the frequency distributions of each independent variable in your hypotheses. If it appears that one has little valid data or was coded with more than 10 categories, substitute another independent variable.
    b. Generate cross-tabulations that show the association between CAPPUN and each of the independent variables. Make sure that CAPPUN is the row variable and that you select "Column Percents."
    c. Does support for capital punishment vary across the categories of any of the independent variables? By how much? Would you conclude that there is an association, as hypothesized, for any pairs of variables?
    d. Might one of the associations you have just identified be spurious because of the effect of a third variable? What might such an extraneous variable be? Look through the variable list and find a variable that might play this role. If you can't think of any possible extraneous variables, or if you didn't find an association in support of any of your hypotheses, try this: Examine the association between CAPPUN and WRKSTAT2. In the next step, control for sex (gender).

       The idea is that there is an association between work status and support for capital punishment that might be spurious because of the effect of sex (gender). Proceed with the following steps:
        i. Select Analyze/Descriptive statistical/Crosstabs.
        ii. In the Crosstabs window, highlight CAPPUN and then click the right arrow to move it into Rows, Move WRKSTAT2 into Columns and SEX into Layer 1 of 1.
        iii. Select Cells/Percentages Column/Continue/OK.

            Is the association between employment status and support for capital punishment affected by gender? Do you conclude that the association between CAPPUN and WRKSTAT2 seems to be spurious because of the effect of SEX?
2. Does the association between support for capital punishment and any of your independent variables vary with social context? Marian Borg (1997) concluded that it did. Test this conclusion by reviewing the association between attitude toward African Americans (HELPBLK) and CAPPUN. Follow the procedures in SPSS Exercise 1d, but click HELPBLK into columns and REGION4 into Layer 1 of 1. (You must first return the variables used previously to the variables list.) Take a while to study this complex three-variable table. Does the association between CAPPUN and HELPBLK vary with region? How would you interpret this finding?
3. Now, how about the influence of an astrological sign on support for capital punishment? Create a cross-tabulation in which ZODIAC is the independent (column) variable and CAPPUN is the dependent (row) variable (with column percents). What do you make of the results?

Developing a Research Proposal

How will you try to establish your hypothesized causal effects (Exhibit 3.10, #9, #10, #11, #16)?

1. Identify at least one hypothesis involving what you expect is a causal relationship. Be sure to specify whether your units of analysis will be individuals or groups.
2. Identify the key variables that should be controlled in your survey design to increase your ability to avoid arriving at a spurious conclusion about the hypothesized causal effect. Draw on relevant

research literature and social theory to identify these variables.

3. Add a longitudinal component to your research design. Explain why you decided to use this particular longitudinal design.
4. Review the criteria for establishing a nomothetic causal effect and discuss your ability to satisfy each one. If you have decided to adopt an idiographic causal approach, explain your rationale.

Section III Basic Social Research Designs

Chapter 7 Experiments

- Research That Matters, Questions That Count
- History of Experimentation
- Careers and Research
- True Experiments
  - Experimental and Comparison Groups
  - Pretest and Posttest Measures
  - Randomization
- Research in the News: Airbnb Hosts and the Disabled
  - Limitations of True Experimental Designs
  - Summary: Causality in True Experiments
- Quasi-Experiments
  - Nonequivalent Control Group Designs
    - *Aggregate Matching*
    - *Individual Matching*
  - Ex Post Facto Control Group Designs
  - Before-and-After Designs
  - Summary: Causality in Quasi-Experiments
- Validity in Experiments
  - Causal (Internal) Validity
    - *Sources of Internal Invalidity Reduced by a Comparison Group*
    - *Sources of Internal Invalidity Reduced by Randomization*
    - *Sources of Internal Invalidity That Require Attention While the Experiment Is in Progress*
  - Generalizability
    - *Sample Generalizability*
    - *Factorial Surveys*
    - *External Validity*
    - *Interaction of Testing and Treatment*
- Ethical Issues in Experimental Research
  - Deception
  - Selective Distribution of Benefits
- Conclusions

Do you think that taking a part-time job below your skill level while you are searching for a full-time professional position after you graduate will harm your prospects for success? Could it make it appear that you lacked sufficient commitment, or failed to make a good impression on other potential employers? Are you sure? Maybe it would demonstrate a high level of persistence in the face of adversity. And do you think it would matter if you are a woman, who is much more likely than a man to have sought a part-time employment in order to have time for child care? Do job hunters who accept part-time or temporary employment tend to have less education or poorer skills? As you learned in Chapter 6, using an experimental design allows researchers to identify the unique effect of a hypothesized causal influence.

Research That Matters, Questions That Count

Millions of workers are employed in temporary and part-time jobs, and in positions that don't utilize their skills, education, or experience. Does a history of such nonstandard, mismatched positions penalize these workers when they search for full-time positions that reflect their training and preparation? Are these workers seen as failing to live up to the "ideal worker" standard, or to have been on the "mommy track"? To put it bluntly, is it a mistake to work as a barista in a coffee shop after you graduate, just to have a job?

Professor David S. Pedulla at the University of Texas at Austin designed an experiment to answer these questions. In 2012 and 2013, 2,420 applications were submitted to 1,210 job openings posted online, and Pedulla tracked the employer callback rate (requests to interview the applicant). The applicant résumés were written so that equal numbers reflected a previous year of work in a full-time standard job, a part-time job, a job through a temporary employment agency, a job below the individual's skill level, or a period of unemployment. The applicants also appeared to differ in gender (using gendered names, such as "Matthew Stevens" or "Emily Stevens").

1. What made this an experiment was that each job opening was randomly assigned to receive two applications with different employment histories, on different days, from either a male or a female applicant. Why do you think this "randomization" procedure was used?
2. In addition to this "field experiment," Pedulla conducted a "survey experiment" to investigate the reasons why nonstandard and mismatched employment histories shape employers' evaluations of job applicants. For this study, 903 individuals who make hiring decisions for U.S. firms answered questions in an online survey about their evaluation of two hypothetical applicants who differed in their employment histories and were either male or female. Does the "field experiment" or "survey experiment" approach seem to you to be more likely to indicate how nonstandard and mismatched job applicants are actually treated in the job market?

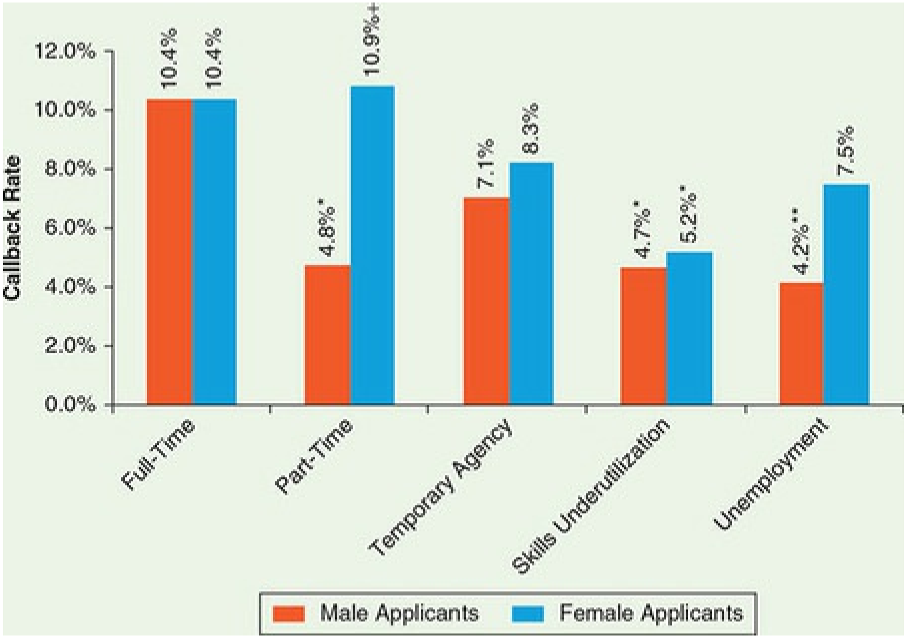
In this chapter, you will learn about procedures for designing experiments and quasi-experiments to test hypotheses. By the end of the chapter, you will understand why an experimental design is the strongest design for establishing causal effects. In addition, you will be able to identify particular features of the design that strengthen its ability to identify causal effects and you will appreciate some of the difficulties of using an experimental design. After you finish the chapter, test yourself by reading the 2016 *American Sociological Review* article by Pedulla at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 7 at edge.sagepub.com/schutt9e.

Pedulla, David S. 2016. "Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories." *American Sociological Review* 81(2):262–289.

Experiments are often the method of choice when the research question is about the effect of a treatment or some other variable whose values can be manipulated by the researcher. As you can see in Exhibit 7.1, Pedulla's experimental design allowed him to identify an adverse effect of having had a period of part-time work or unemployment on the chances of employers calling back male but not female job applicants, and that having worked in a job in which their skills were underutilized harmed both. Social psychologists have often used experiments to study bias and other attitudes that are hard to isolate from other influences.

M. Lyn Exum's (2002) research about the impact of intoxication and anger on aggressive intentions was another example (described in Chapter 6). You will read about other fascinating experimental studies in this chapter, but bear in mind that many research questions posed by sociologists, political scientists, and other social scientists do not lend themselves to investigation with experimental designs.

Exhibit 7.1 Callback Rates, by Employment History and Gender



Source: Pedulla 2016:273.

This chapter examines experimental methodology in greater detail. First, you will read about the history of experimentation in the social sciences and then about different types of experimental design (both true experiments and quasi-experiments). Next, you will learn about the ability of particular designs to establish causally valid conclusions and to achieve generalizable results. Finally, you will consider ethical issues that should be given particular attention when considering experimental research.

History of Experimentation

Experimentation has always been a key method of scientific research. A famous experiment, perhaps the first, was conducted by Archimedes in 230 BC when he tested the way in which levers balance weights. In 1607, Galileo conducted experiments by dropping weights to test the theory of falling bodies. By the mid-1800s, the use of experiments in the rapidly growing fields of scientific research was well established (Willer and Walker 2007:Par. 2 in "A Brief History of Two Kinds of Experiments"). Both examples illustrate the central feature of experiments (Willer and Walker 2007):

> an inquiry for which the investigator controls the phenomena of interest and sets the conditions under which they are observed and measured. (Par. 2 in "Experiments: A Definition")

Successful experiments control the conditions in which a hypothesis is tested and measure variables precisely. Conducting the research in a laboratory setting—often a specially designed room in a university—can permit exacting control over every detail. Experimental researchers thus remove sources of variability that are not relevant to the specific hypothesis being tested (Willer and Walker 2007:"Comparing Experiments and Nonexperimental Studies").

Careers and Research

Sruthi Chandrasekaran, Senior Research Associate

Sruthi Chandrasekaran is a senior research associate at J-PAL—the Abdul Latif Jameel Poverty Action Lab that was established at the Massachusetts Institute of Technology but has become a global network of researchers who seek to reduce poverty by ensuring that policy is informed by scientific evidence. J-PAL emphasizes the use of randomized controlled trials to evaluate the impact of social policies.

Chandrasekaran has completed a 5-year integrated master's in economics at the Indian Institute of Technology (IIT) Madras and an MSc in comparative social policy at the University of Oxford. Her most recent project tests the value of performance-based incentives on improving tuberculosis (TB) reduction efforts by health workers in North Indian slums.

Chandrasekaran's academic training in economics and social policy provided strong qualitative and quantitative research tools, but her interest in having an impact on societal development led to her career. As a field-based researcher, she meets with communities, listens to their perspectives, and proposes interventions. She then takes the lead in ensuring that the intervention follows the study design to the dot, the data collection tools elicit quality responses in an unbiased manner, the survey data are of the highest quality, the cleaning of the data is coherent and methodical, and the analysis is rigorous. Because study results are published in leading academic journals and the policy lessons are disseminated to key stakeholders, it is crucial that the research is well designed and the quality of the data is impeccable.

Chandrasekaran's research training helps her examine issues in an objective manner, develop a logical framework to investigate issues in detail, and understand the story behind the data. She also strives to affect policy design and implementation by sharing what she has learned in the field. Working with data collected about real problems helps make these tasks interesting, exciting, and rewarding.

Chandrasekaran offers some heartfelt advice for students interested in a career involving doing research or using research results:

> Researchers need the ability to study an aspect of a social problem in great detail as well as the flexibility to step back and look at the bigger picture. Consciously training to don both hats is very helpful. The ability to understand field realities is crucial to designing a research question that is grounded as well as one that is useful for policy analysis. Research can at times be painstakingly slow and frustrating, so patience and single-minded focus on the end goal can help one through the tough times. Being aware of competing methodologies and research studies in relevant fields can also be quite useful in understanding the advantages and pitfalls in your own research. If you are inspired to take up research, make sure you choose a field close to your heart since this will be personally and professionally rewarding. If you are unsure, take up an internship or a short-term project to see how much you may enjoy it.

Experimental research by sociologists, social psychologists, and psychologists as well as other social scientists has made important contributions to understanding the social world since the middle years of the 20th century. Laboratory experiments by Harvard's Solomon Asch (1958) showed that individuals would yield to social pressure in their own judgments about the length of lines—even when they knew that their perceptions were correct. Experiments by Joseph Berger, Bernard Cohen, and Morris Zelditch Jr. (1972) made it clear that students quickly deferred in conversation to other students of higher status. John Bargh, Mark Chen, and Lara Burrows (1996) found that having students read words that included some polite, neutral, or rude terms had a "priming effect" that led them immediately after to behave in a corresponding manner. Responding to concerns about the effect of economic hardship on crime, Richard Berk, Kenneth Lenihan, and Peter Rossi (1980) conducted an experimental test "in the field" of the impact of providing financial subsidies to released convicts for 1 to 2 years. (It had no impact on their likelihood of reoffending.)

True Experiments

The value of experimental methods is clearest when we focus on what are termed *true experiments.* True experiments must have at least three features:

1. Two groups (in the simplest case, an experimental and a control group)
2. Variation in the independent variable before assessment of change in the dependent variable
3. Random assignment to the two (or more) comparison groups

The combination of these features permits us to have much greater confidence in the validity of causal conclusions than is possible in other research designs. As you learned in Chapter 6, two more features further enhance our confidence in the validity of an experiment's findings:

1. Identification of the causal mechanism
2. Control over the context of an experiment

You will learn more about each of these key features of experimental design as you review three different experimental studies about social processes. I use simple diagrams to help describe and compare the experiments' designs. These diagrams also show at a glance how well suited any experiment is to identifying causal relationships, by indicating whether the experiment has a comparison group, a pretest and a posttest, and randomization.

Experimental and Comparison Groups

True experiments must have at least one experimental group (subjects who receive some treatment) and at least one comparison group (subjects with whom the experimental group can be compared). The comparison group differs from the experimental group in one or more independent variables, whose effects are being tested. In other words, variation in the independent variable determines the difference between the experimental and comparison groups.

In many experiments, the independent variable indicates the presence or absence of something, such as receiving a treatment program or not receiving it. In these experiments, the comparison group, consisting of the subjects who do not receive the treatment, is termed a control group. You learned in Chapter 6 that an experiment can have more than two groups. There can be several treatment groups, corresponding to different values of the independent variable, and several comparison groups, including a control group that receives no treatment.

True experiment: An experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation; outcomes are measured in a posttest.

Experimental group: In an experiment, the group of subjects that receives the treatment or experimental manipulation.

Comparison group: In an experiment, a group that has been exposed to a different treatment (or value of the independent variable) than the experimental group.

Control group: A comparison group that receives no treatment.

Alexander Czopp, Margo Monteith, and Aimee Mark (2006) used a control group in an experiment about reducing bias through interpersonal confrontation. For this experiment, Czopp et al. recruited 111 white students from introductory psychology classes and had them complete a long survey that included an “Attitude Toward Blacks” (ATB) scale. At the time of the experiment, individual students came to a laboratory, sat before a computer, and were told that they would be working with a student in another room to complete a task. All interaction with this “other subject” would be through the computer. Unknown to the student recruits, the other student was actually the experimenter. Also without their knowledge, the student recruits were assigned randomly to one of three conditions in the experiment.

The students first answered some “getting to know you” questions from the “other student." The "other student" did not mention his or her gender, but did mention that he was white. After this brief Q&A, the student subjects were shown images of various people, of different races, and asked to write a comment about them to the "other student." The "other student" did the same thing, according to a standard script, and had an opportunity to give some feedback about these comments. Here is the type of feedback that students in the Racial Confrontation condition soon received (the particular words varied according to what the student subject had actually written):

> i thought some of your answers seemed a little offensive. the Black guy wandering the streets could be a lost tourist and the Black woman could work for the government. people shouldn't use stereotypes, you know? (Czopp et al. 2006:795)

Students in the Nonracial Confrontation condition received a response meant to provoke but without any racial overtones, like, "I thought some of our answers seemed a little goofy. 'A traveler spends time in airports?' 'A librarian has a lot of books?' Couldn't you think of anything better than that?" Students in the No-Confrontation Control condition were told only, "I thought you typed fast. Good job." Shortly after this, the subjects completed the ATB again as well as other measures of their attitudes.

The study results indicated that the interpersonal confrontations were effective in curbing stereotypic responding, whether the confronter was white or black (see Exhibit 7.2). Students who were "confronted" in this situation also indicated a more negative conception of themselves after the confrontation.

Pretest and Posttest Measures

All true experiments have a posttest—that is, measurement of the outcome in both groups after the experimental group has received the treatment. Many true experiments also have pretests that measure the dependent variable before the experimental intervention. A pretest is exactly the same as a posttest, just administered at a different time. Pretest scores permit a direct measure of how much the experimental and comparison groups changed over time, such as the change in stereotypic responding among the students that Czopp et al. studied.

Pretest scores also allow the researcher to verify that randomization was successful (that chance factors did not lead to an initial difference between the groups). In addition, by identifying subjects' initial scores on the dependent variable, a pretest provides a more complete picture of the conditions in which the intervention had (or didn't have) an effect (Mohr 1992:46–48). A randomized experimental design with a pretest and posttest is termed a randomized comparative change design or a *pretest–posttest control group design.*

An experiment may have multiple posttests and perhaps even multiple pretests. Multiple posttests can identify just when the treatment has its effect and for how long. This is particularly important when treatments are delivered over time (Rossi and Freeman 1989:289–290).

Posttest: In experimental research, the measurement of an outcome (dependent) variable after an experimental intervention or after a presumed independent variable has changed for some other reason.

Pretest: In experimental research, the measurement of an outcome (dependent) variable before an experimental intervention or change in a presumed independent variable for some other reason. The pretest is exactly the same "test" as the posttest, but it is administered at a different time.

Randomized comparative change design: The classic true experimental design in which subjects are assigned randomly to two groups; both these groups receive a pretest, then one group receives the experimental intervention, and then both groups receive a posttest. Also known as a *pretest–posttest control group design.*

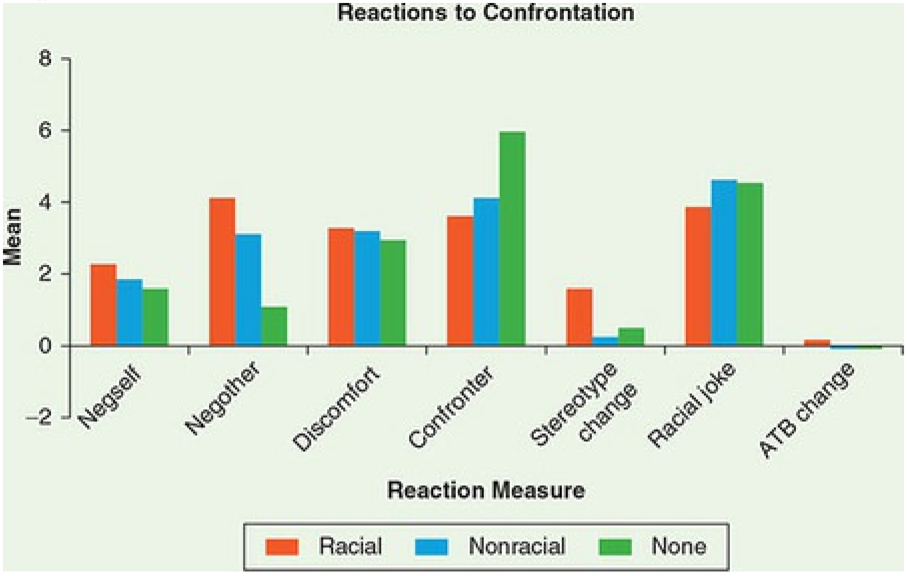
Strictly speaking, however, a true experiment does not require a pretest. When researchers use random assignment to the experimental and comparison groups, the groups' initial scores on the dependent variable and on all other variables are likely to be similar. Any difference in outcome between the experimental and comparison groups is therefore likely to result from the intervention (or to other processes occurring during the experiment), and the likelihood of a difference just on the basis of chance can be calculated. This is fortunate, because the dependent variable in some experiments cannot be measured in a pretest. For example, Czopp et al. (2006:797) measured the attitudes of their student subjects toward the "other student" who had confronted them. They weren't able to measure this attitude until after the interaction in which they manipulated the confrontation. Thus, Exhibit 7.2 includes some measures that represent change from the pretest to the posttest (ATB scores) and some that represent only scores at the posttest. Exhibit 7.3 diagrams the Czopp study. The labels indicate that the *pretest–posttest control group design* was used with the ATB measure, whereas the other measures in the posttest were used in a *posttest-only control group design,* also called the randomized comparative posttest design. You'll also learn later in this chapter that there can be a disadvantage to having a pretest, even when it is possible to do so: The act of taking the pretest can itself cause subjects to change.

Randomized comparative posttest design: A true experimental design in which subjects are assigned randomly to two groups—one group then receives the experimental intervention and both groups receive a posttest; there is no pretest. Also known as a *posttest-only control group design.*

Randomization

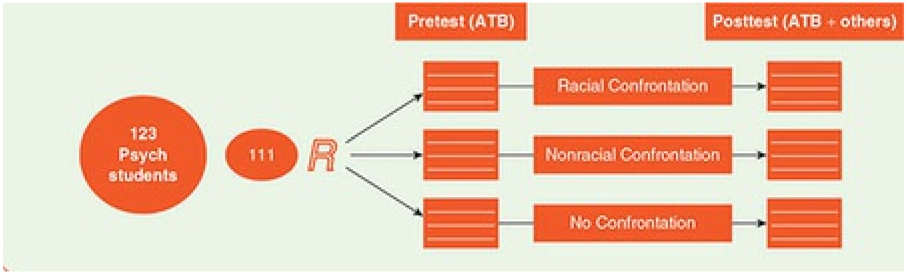
Randomization, or random assignment, is what makes the comparison group in a true experiment such as Czopp et al.'s such a powerful tool for identifying the effects of the treatment. A randomized comparison group can provide a good estimate of the counterfactual—the outcome that would have occurred if the subjects who were exposed to the treatment actually had not been exposed but otherwise had had the same experiences (Mohr 1992:3; Rossi and Freeman 1989:229). A researcher cannot determine for sure what the unique effects of a treatment are if the comparison group differs from the experimental group in any way other than not receiving the treatment.

Exhibit 7.2 Students' Reactions to Confrontation Conditions in Czopp et al. Experiment (scores not standardized)



Source: Czopp, Alexander M., Margo J. Monteith, and Aimee Y. Mark. 2006. "Standing Up for a Change: Reducing Bias Through Interpersonal Confrontation." Journal of Personality and Social Psychology 90:784–803.

Exhibit 7.3 Diagram of Confrontation Experiment 2


Pretest (ATB)
Posttest (ATB + others)
123 Psych students
111
R
Racial Confrontation
Nonracial Confrontation
No Confrontation

Source: Based on Czopp et al. (2006:795–796).

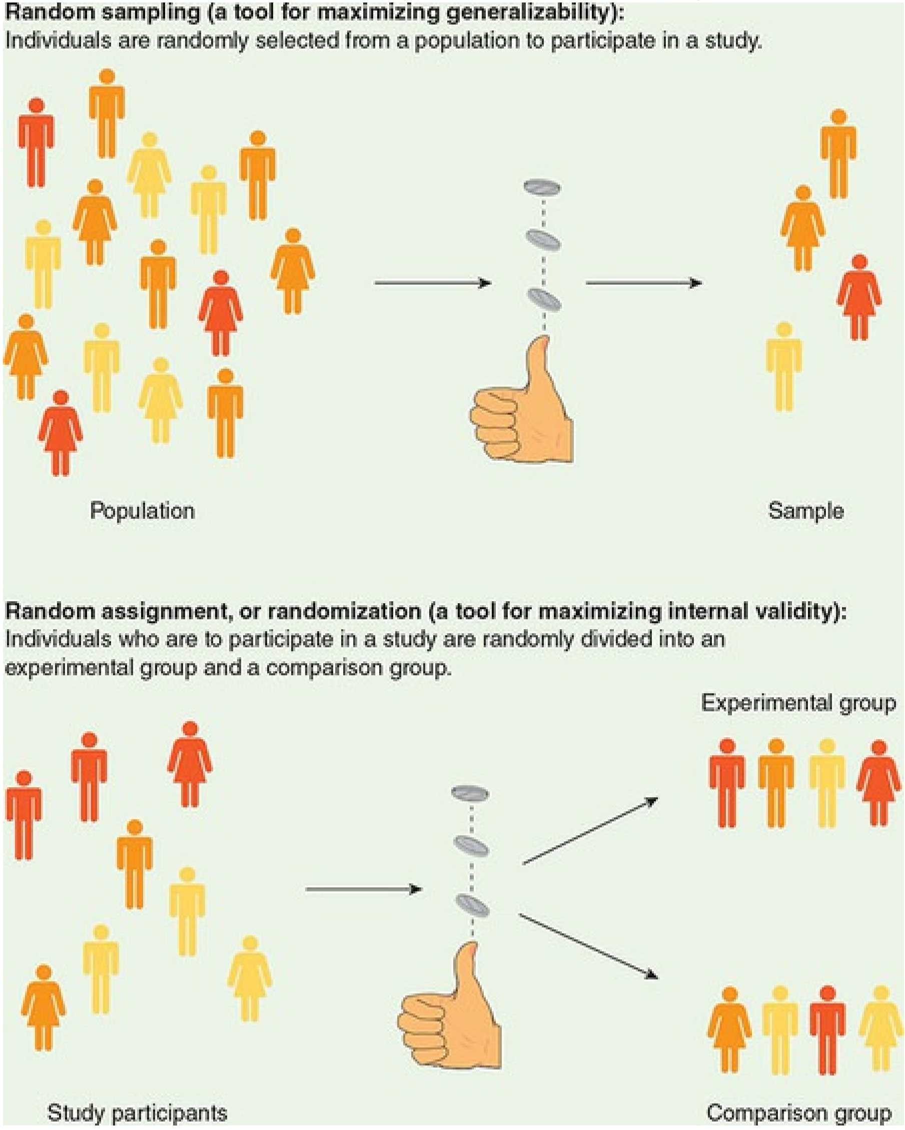
Assigning subjects randomly to the experimental and comparison groups ensures that systematic bias does not affect the assignment of subjects to groups. Of course, random assignment cannot guarantee that the groups are perfectly identical at the start of the experiment. Randomization removes bias from the assignment process, but only by relying on chance, which itself can result in some intergroup differences (Bloom 2008:116). Fortunately, researchers can use statistical methods to determine the odds of ending up with groups that differ very much on the basis of chance, and these odds are low even for groups of moderate size. The larger the group, the less likely it is that even modest differences will occur on the basis of chance and the more possible it becomes to draw conclusions about causal effects from relatively small differences in the outcome.

Note that the random assignment of subjects to experimental and comparison groups is not the same as random sampling of individuals from some larger population (see Exhibit 7.4). In fact, random assignment (randomization) does not help at all to ensure that the research subjects are representative of some larger population; instead, representativeness is the goal of random sampling. What random assignment does—create two (or more) equivalent groups—is useful for maximizing the likelihood of internal validity, not generalizability (Bloom 2008:116).

Matching is another procedure used to equate experimental and comparison groups, but by itself it is a poor substitute for randomization. Matching of individuals in a treatment group with those in a comparison group might involve pairing persons on the basis of similarity of gender, age, year in school, or some other characteristic. The basic problem is that, as a practical matter, individuals can be matched on only a few characteristics; as a result, unmatched differences between the experimental and comparison groups may still influence outcomes. When matching is used as a substitute for random assignment, the research becomes quasi-experimental instead of being a true experiment. However, matching combined with randomization, also called *blocking,* can reduce the possibility of differences resulting from chance (Bloom 2008:124). For example, if individuals are matched in gender and age, and then the members of each matched pair are assigned randomly to the experimental and comparison groups, the possibility of outcome differences because of differences in the gender and age composition of the groups is eliminated (see Exhibit 7.5).

Matching: A procedure for equating the characteristics of individuals in different comparison groups in an experiment. Matching can be done on either an individual or an aggregate basis. For individual matching, individuals who are similar in key characteristics are paired before assignment, and then the two members of each pair are assigned to the two groups. For aggregate matching, also termed *blocking,* groups that are chosen for comparison are similar in the distribution of key characteristics.

Exhibit 7.4 Random Sampling Versus Random Assignment

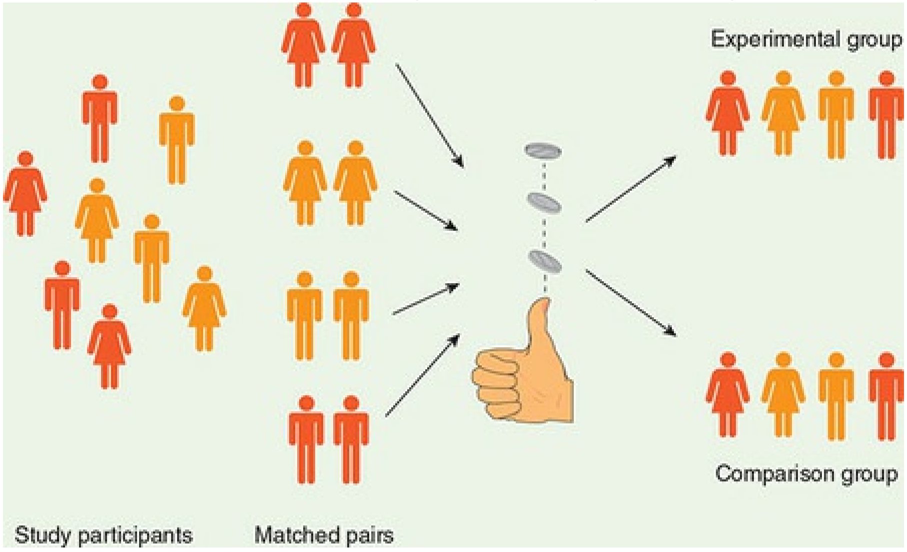

Random sampling (a tool for maximizing generalizability):
Individuals are randomly selected from a population to participate in a study.
Population
Sample
Random assignment, or randomization (a tool for maximizing internal validity):
Individuals who are to participate in a study are randomly divided into an experimental group and a comparison group.
Experimental group
Study participants
Comparison group

The Solomon four-group design is a true experimental design that combines a randomized comparative change design (the pretest–posttest control group design) with the randomized comparative posttest design (posttest-only control group design). This design allows comparison of the effect of the independent variable on groups that had a pretest with its effect on groups that have not had a pretest (see Exhibit 7.11, later in this chapter). Whenever there is reason to think that taking the pretest may itself influence how participants react to the treatment, the Solomon four-group design should be considered. If the pretest had such an effect, the difference in outcome scores between the experimental and comparison groups will be different for subjects who took the pretest (the pretest–posttest design) compared with those who did not (the posttest-only design).

Solomon four-group design: A type of experimental design that combines a randomized pretest–

posttest control group design with a randomized posttest-only design, resulting in two experimental groups and two comparison groups.

Exhibit 7.5 Experimental Design Combining Matching and Random Assignment


Experimental group
Comparison group
Study participants
Matched pairs

Limitations of True Experimental Designs

The distinguishing features of true experiments—experimental and comparison groups, pretests (which are not always used) and posttests, and randomization—can be implemented most easily in laboratory settings with participants such as students who are available for such experiments. For this reason, true experimental designs are used most often in social psychological experiments that focus on research questions about reactions to conditions that can easily be created in laboratories on college campuses. However, this focus on college students in laboratory settings raises the question of whether findings can be generalized to other populations and settings. This problem of generalizability is the biggest limitation of true experimental designs and so I will return to it later in the chapter. You will read examples of experiments conducted outside of the laboratory in Chapter 13, on evaluation and policy research.

The potential for the pretest to influence participants' reactions to the experimental treatment is also a limitation of experimental designs, but it is one that can be solved by researchers who have sufficient funds and time to double the number of participants in the experiment and use a Solomon four-group design.

The criteria for true experimental designs also do not help researchers identify the mechanisms by which treatments have their effects. In fact, this question of causal mechanisms often is not addressed in experimental research. The hypothesis test itself does not require any analysis of mechanism, and if the experiment was conducted under carefully controlled conditions during a limited time span, the causal effect (if any) may seem to be quite direct. But attention to causal mechanisms can augment experimental findings. For example, evaluation researchers often focus attention on the mechanisms by which a social program has its effect (Mohr 1992:25–27; Scriven 1972a). The goal is to measure the intermediate steps that lead to the change that is the program's primary focus.

True experimental designs also do not guarantee that the researcher has been able to maintain control over the conditions to which subjects are exposed after they are assigned to the experimental and comparison groups. If these conditions begin to differ, the variation between the experimental and comparison groups will not be what was intended. Such unintended variation is often not much of a problem in laboratory experiments, where the researcher has almost complete control over the conditions (and can ensure that these conditions are nearly identical for both groups). But control over conditions can become a big concern for field experiments, experimental studies that are conducted in the field, in real-world settings. Pedulla's (2016) experimental study of employer responses to different types of job applicants was conducted "in the field" and didn't require any change in the usual procedures with job applicants. By contrast, in Sherman and Berk's (1984) field experiment about the police response to domestic violence (see Chapter 2), police officers did not adhere consistently to the requirement that they arrest suspects based on the random assignment protocol. As a result of this concern, the subsequent replications of the experiment in other cities reduced the discretion of individual police officers and used a more centralized procedure for assigning cases to the treatment and comparison conditions.

Field experiment: A study using an experimental design that is conducted in a real-world setting.

Summary: Causality in True Experiments

The study by Czopp et al. (2006) was a true experiment because it had at least one experimental and one comparison group to which subjects were randomly assigned. The researchers also compared variation in the dependent variables after variation in the independent variable (confrontation condition). Czopp et al. had a pretest score for only one of the variables they measured in the posttest, but you have learned that a pretest is not required in a true experiment. Brad Bushman, Roy Baumeister, and Angela Stack's (1999) experimental study of catharsis (Chapter 6) did not have a pretest, nor did Sherman and Berk's (1984) experimental study of the police response to domestic violence (Chapter 2).

Let's examine how well true experiments meet the criteria for identifying a nomothetic cause (introduced in Chapter 6):

Association between the hypothesized independent and dependent variables.

As you have seen, experiments can provide unambiguous evidence of association by comparing the distribution of the dependent variable (or its average value) between the experimental and comparison groups.

Time order of effects of one variable on the others.

Unquestionably, arrest for spouse abuse preceded recidivism in the Sherman and Berk (1984) study (described in Chapter 2), and the "confrontations" in the Czopp et al. (2006) study preceded the differential changes in prejudicial attitudes between the experimental and comparison groups. In true experiments, randomization to the experimental and comparison groups equates the groups at the start of the experiment, so time order can be established by comparing posttest scores between the groups. However, experimental researchers include a pretest when possible so that equivalence of the groups at baseline can be confirmed and the amount of change can be compared between the experimental and comparison groups.

Nonspurious relationships between variables.

Nonspuriousness is difficult—some would say impossible—to establish in nonexperimental designs. The random assignment of subjects to experimental and comparison groups is what makes true experiments such powerful designs for testing causal hypotheses. Randomization controls for the host of possible extraneous influences that can create misleading, spurious relationships in both experimental and nonexperimental data. If we determine that a design has used randomization successfully, we can be much more confident in the resulting causal conclusions.

Mechanism that creates the causal effect.

The features of a true experiment do not in themselves allow identification of causal mechanisms; as a result, there can be some ambiguity about how the independent variable influenced the dependent variable and the resulting causal conclusions (Bloom 2008:128). However, Czopp et al. (2006:798) investigated possible mechanisms linking confrontation to change in prejudicial attitudes in their experiment. One finding from this investigation was that the confrontations led to a more negative self-appraisal, which in turn led to decreased expression of prejudice.

Context in which change occurs.

Control over conditions is more feasible in many experimental designs than it is in nonexperimental designs. Czopp et al. (2006) allowed their student subjects to communicate with the “other student” collaborator only through a computer to maintain control over conditions. The researchers didn’t want the student subjects to notice something about the “other student” that might not have to do with their manipulation about confrontation. In another version of the experiment, Czopp and colleagues compared the responses of student subjects with “other students” who were said to be black and white (and found that the race of the confederate did not matter) (pp. 791–794). Bear in mind that it is often difficult to control conditions in experiments conducted outside of a laboratory setting; later in this chapter, you will see how the lack of control over experimental conditions can threaten internal validity.

Quasi-Experiments

Often, testing a hypothesis with a true experimental design is not feasible with the desired participants and in the desired setting. Such a test may be too costly or take too long to carry out, it may not be ethical to randomly assign subjects to the different conditions, or it may be too late to do so. In these situations, researchers may instead use *quasi-experimental* designs that retain several components of experimental design but do not randomly assign participants to different conditions.

A quasi-experimental design is one in which the comparison group is predetermined to be comparable with the treatment group in critical ways, such as being eligible for the same services or being in the same school cohort (Rossi and Freeman 1989:313). These research designs are *quasi-*experimental because subjects are not randomly assigned to the comparison and experimental groups. As a result, we cannot be as confident in the comparability of the groups as in true experimental designs. Nonetheless, to term a research design quasi-experimental, we have to be sure that the comparison groups meet specific criteria that help lessen the possibility of preexisting differences between groups.

I discuss here the two major types of quasi-experimental designs—nonequivalent control group designs and before-and-after designs—as well as a nonexperimental design that can be very similar to nonequivalent control group designs (other types can be found in Cook and Campbell 1979; Mohr 1992):

1. Nonequivalent control group designs have experimental and comparison groups that are designated before the treatment occurs and are not created by random assignment.
2. Ex post facto control group designs have experimental and comparison groups that are not designated before the treatment occurs and are not created by random assignment, so that participants may select themselves to be in a group. This ability to choose the desired type of group actually makes this design nonexperimental rather than quasi-experimental, but it is often confused with the nonequivalent control group design.
3. Before-and-after designs have a pretest and posttest but no comparison group. In other words, the subjects exposed to the treatment serve, at an earlier time, as their own controls.

Quasi-experimental design: A research design in which there is a comparison group that is comparable to the experimental group in critical ways, but subjects are not randomly assigned to the comparison and experimental groups.

Nonequivalent control group design: A quasi-experimental design in which experimental and comparison groups are designated before the treatment occurs but are not created by random assignment.

Ex post facto control group design: A nonexperimental design in which comparison groups are selected after the treatment, program, or other variation in the independent variable has occurred, but when the participants were able to choose the group in which they participated. Often confused with a quasi-experimental design.

Before-and-after design: A quasi-experimental design consisting of several before-after comparisons involving the same variables but no comparison group.

Nonequivalent Control Group Designs

The nonequivalent control group design is the most common type of quasi-experimental design, also called a *differences-in-differences* design. In it, a comparison group is selected to be as comparable as possible to the treatment group. Two selection methods can be used: aggregate matching and individual matching.

In the News Research in the News: Airbnb Hosts and the Disabled

For Further Thought?

Rates of preapproval by Airbnb hosts dropped for travelers who said they had a disability, according to a 2016 study. Lisa Schur at the Rutgers School of Management and Labor Relations and others e-mailed more than 3,800 Airbnb lodging requests and found that 75% were granted preapproval—if no disability was mentioned. The preapproval rate dropped to 61% for those who said they had dwarfism, 50% of those who said they were blind, 43% for those with cerebral palsy, and 25% for those with spinal cord injuries.

In their article about the Airbnb study, Niraj Chokshi and Katie Brenner also noted that Airbnb has instituted a new non-discrimination policy and that most Airbnb reservations are made without a preapproval requirement.

1. Would you classify this study as an experiment? What features make it (or would make it) experimental?
2. What causal mechanisms might explain the effect of having a disability that Schur identified (some are identified in the news article)? How would you design a study to investigate these mechanisms?

News source: Chokshi, Niraj and Katie Brenner. 2017. “Disabled Travelers Are More Likely to Be Rejected by Airbnb Hosts, a Study Finds.” *The New York Times,* June 3, p. B5.

Aggregate Matching

Once research moves outside of laboratories on college campuses and samples of available students, it becomes much harder to control what people do and what type of experiences they will have. When random assignment is not possible, an alternative approach to testing the effect of some treatment or other experience can be to find a comparison group that matches the treatment group in many ways but differs in exposure to the treatment. For example, a sociologist who hypothesizes that the experience of a disaster increases social solidarity might identify two towns that are similar in population characteristics but where one experienced an unexpected disaster. If the populations in the two towns have similar distributions on key variables such as age, gender, income, and so on, but they differ after the disaster in their feelings of social solidarity, the sociologist might conclude that a higher level of solidarity in the affected town resulted from the disaster. However, it is important in a nonequivalent control group design that individuals have not been able to choose whether to join the group that had one experience or the other. If many people moved from the stricken town to the unaffected town right after the disaster, higher levels of post-disaster solidarity among residents in the stricken town could reflect the departure of people with feelings of low solidarity from that town, rather than the effect of the disaster on increasing feelings of solidarity.

Siw Tone Innstrand, Geir Arild Espries, and Reidar Mykletun (2004) used a nonequivalent control group design with aggregate matching to study the effect of a new program in Norway to reduce stress among staff working with people with intellectual disabilities. The researchers chose two Norwegian municipalities that offered the same type of community residential care and used similar staff. The two municipalities were in different locations and had no formal communication with each other.

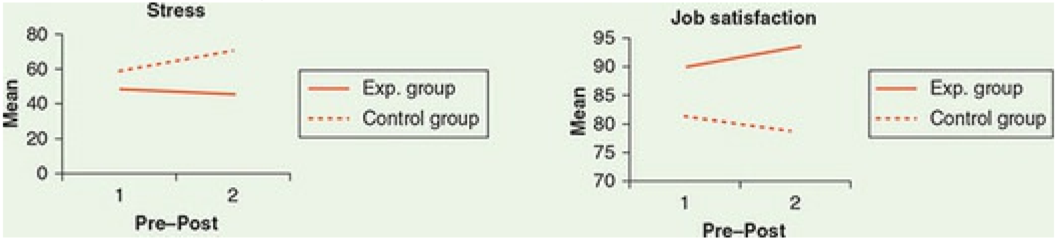
When their research began, Innstrand and colleagues asked staff in both municipalities to complete an anonymous survey that measured their stress, job satisfaction, and other outcomes. This was their pretest. The researchers then implemented the stress reduction program in one municipality. This was their experimental treatment. They distributed their survey again 10 months later. This was their posttest. Two of their primary findings appear in Exhibit 7.6: Levels of stress declined in the experimental group and increased in the control group, and job satisfaction increased in the experimental group and declined in the control group. It seemed that the program had at least some of the effects that Innstrand and colleagues predicted.

Nonequivalent control group designs based on aggregate matching should always be considered when the goal is to compare effects of treatments or other experiences and a true experiment is not feasible. However, simply comparing outcome measures in two groups that offer different programs is not in itself a quasi-experiment. If individuals can choose which group to join, partly on the basis of the program offered, then the groups will differ in preference for the treatment as well as in having had the treatment. When such selection bias is possible, the design is nonexperimental rather than quasi-experimental. More generally, the validity of this design depends on the adequacy of matching of the comparison group with the treatment group (Cook and Wong 2008:151).

Individual Matching

In individual matching, individual cases in the treatment group are matched with similar individuals in the comparison group. In some situations, this can create a comparison group that is very similar to the experimental group, as when children in Head Start were matched with their siblings to estimate the effect of participation in Head Start (Currie and Thomas 1995:341). However, in many studies, it is not possible to match in this way on the most important variables. Simply matching on the basis of such readily available characteristics as gender or age will be of little value if these sociodemographic characteristics are not related to propensity to participate in the treatment group or extent of response to treatment (Cook and Wong 2008:153).

Exhibit 7.6 Findings of the Innstrand Quasi-Experimental Study



Source: Innstrand, Siw Tone, Geir Arild Espries, and Reidar Mykletun. 2004. "Job Stress, Burnout and Job Satisfaction: An Intervention Study of Staff Working With People With Intellectual Disabilities." Journal of Applied Research in Intellectual Disabilities 17:119–126. Copyright © 2004, John Wiley & Sons. Reprinted with permission.

Variables chosen for matching affect the quality of a nonequivalent comparison group design (Cook and Wong 2008:154–155). Matching with comparison groups comprising twins, siblings, members of the same organization, or others who are very similar will often be the best choice.

The quality of a nonequivalent comparison group design can also be improved by inclusion of several other features (Cook and Wong 2008:154–155):

- A pretest in both groups, so that the starting point for the groups can be compared
- Identical procedures to measure outcomes in the two groups (same measure, same time of assessment)
- Several outcome measures reflecting the same overall causal hypothesis
- Investigating the process of selection into the two groups and controlling for the elements of this process that can be measured

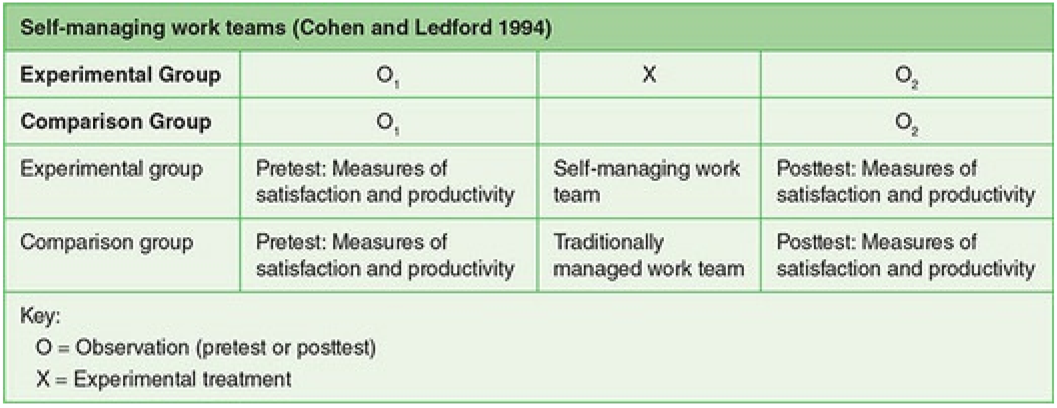
Quasi-experimental designs with these features can result in estimates of effects that are very similar to those obtained with a randomized design (Cook and Wong 2008:159).

Ex Post Facto Control Group Designs

The ex post facto control group design is similar to the nonequivalent control group design and is often confused with it, but it does not meet as well the criteria for quasi-experimental designs. This design has experimental and comparison groups that are not created by random assignment, but unlike nonequivalent control group designs, individuals may decide themselves whether to enter the "treatment" or control group. As a result, in ex post facto (after the fact) designs, the people who join the treatment group may differ because of what attracted them to the group initially, rather than because of their experience in the group. However, in some studies, we may conclude that the treatment and control groups are so similar at the outset that causal effects can be tested (Rossi and Freeman 1989:343–344).

Susan Cohen and Gerald Ledford (1994) studied the effectiveness of self-managing teams in a telecommunications company with an ex post facto design (see Exhibit 7.7). They compared work teams they rated as self-managing with those they found to be traditionally managed (meaning that a manager was responsible for the team's decisions). Cohen and Ledford found that the self-reported quality of work life was higher in the self-managed groups than in the traditionally managed groups.

Exhibit 7.7 Ex Post Facto Control Group Design



Self-managing work teams (Cohen and Ledford 1994)			
Experimental Group	O_1	X	O_2
Comparison Group	O_1		O_2
Experimental group	Pretest: Measures of satisfaction and productivity	Self-managing work team	Posttest: Measures of satisfaction and productivity
Comparison group	Pretest: Measures of satisfaction and productivity	Traditionally managed work team	Posttest: Measures of satisfaction and productivity

Key:
O = Observation (pretest or posttest)
X = Experimental treatment

Source: Based on Cohen and Ledford (1994).

What distinguishes this study design from a quasi-experimental design like the one Wageman (1995) used to study work teams is the fact that the teams themselves and their managers had some influence on how they were managed. As the researchers noted, "If the groups which were already high performers were the ones selected to be self-managing teams, then the findings could be due to a selection bias rather than any effects of self-management" (Cohen and Ledford 1994:34). Thus, preexisting characteristics of employees and managers or their team composition might have influenced which "treatment" they received, as well as the outcomes achieved. This leaves us less certain about the effect of the treatment itself.

Before-and-After Designs

The common feature of before-and-after designs is the absence of a comparison group. Because all cases are exposed to the experimental treatment, the basis for comparison is provided by comparing the pretreatment with the posttreatment measures. These designs are thus useful for studies of interventions that are experienced by virtually every case in some population, including total coverage programs such as Social Security or studies of the effect of a new management strategy in a single organization. The simplest type of before-and-after design is the fixed-sample panel design, with one pretest and one posttest (see Chapter 6).

David Phillips's (1982) study of the effect of TV soap opera suicides on the number of actual suicides in the United States illustrates a more powerful multiple group before-and-after design. In this design, several before-and-after comparisons are made involving the same variables but with different groups. Phillips identified 13 soap opera suicides in 1977 and then recorded the U.S. suicide rate in the weeks before and following each TV story. Because several suicides occurred in adjacent weeks, the analysis proceeded as if there had been 9 soap opera suicides. In effect, the researcher had 9 different before-and-after studies, one for each suicide story occurring in a unique week. In 8 of these 9 comparisons, deaths from suicide increased from the week before each soap opera suicide to the week after (see Exhibit 7.8).

Another type of before-and-after design involves multiple pretest and posttest observations of the same group. These may be repeated measures panel designs, which include several pretest and posttest observations, and time series designs, which include many (preferably 30 or more) such observations in both pretest and posttest periods. Repeated measures panel designs are stronger than simple before-and-after panel designs because they allow the researcher to study the process by which an intervention or treatment has an impact over time.

In a time series design, the trend in the dependent variable until the date of the intervention or event whose effect is being studied is compared with the trend in the dependent variable after the intervention. A substantial disparity between the preintervention trend and the postintervention trend is evidence that the intervention or event had an impact (Rossi and Freeman 1989:260–261, 358–363).

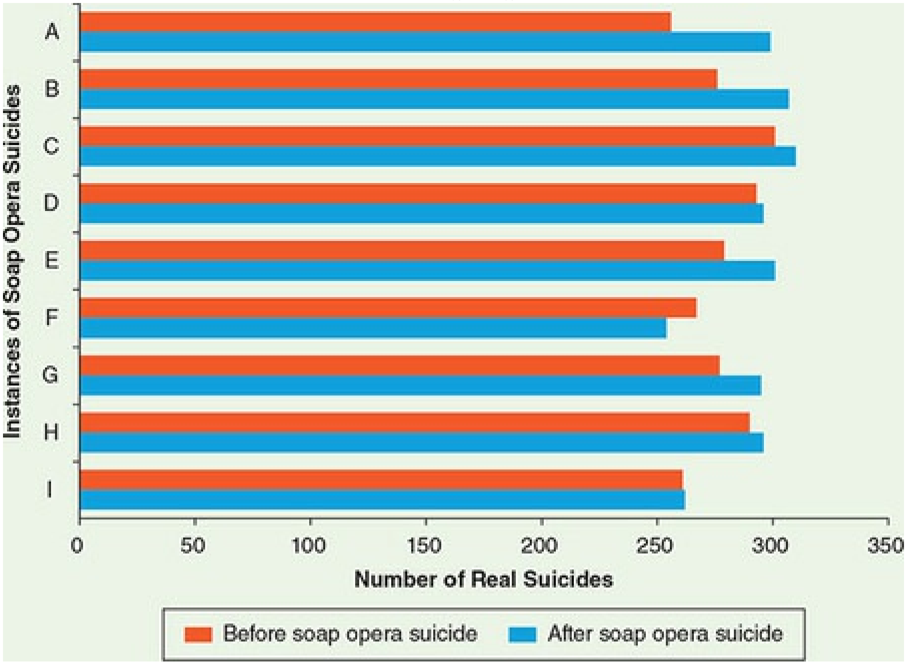
Multiple group before-and-after design: A type of quasi-experimental design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.

Repeated measures panel design: A quasi-experimental design consisting of several pretest and posttest observations of the same group.

Time series design: A quasi-experimental design consisting of many pretest and posttest observations of the same group over an extended period.

Time series designs are particularly useful for studies of the impact of new laws or social programs that affect everyone and that are readily assessed by some ongoing measurement. For example, Paul A. Nakonezny, Rebecca Reddick, and Joseph Lee Rodgers (2004) used a time series design to identify the impact of the Oklahoma City terrorist bombing in April 1995 on the divorce rate in Oklahoma. They hypothesized that people would be more likely to feel a need for support in the aftermath of such a terrifying event and thus be less likely to divorce. Nakonezny et al. first calculated the average rate of change in divorce rates in Oklahoma's 77 counties in the 10 years before the bombing and then projected these rates forward to the 5 years after the bombing. As they hypothesized, they found that the actual divorce rate in the first years after the bombing was lower than the prebombing trend would have predicted, but this effect diminished to nothing by the year 2000 (see Exhibit 7.9).

Exhibit 7.8 Real Suicides and Soap Opera Suicides



Source: Adapted from Phillips (1982):1347. Reprinted with permission from the University of Chicago Press.

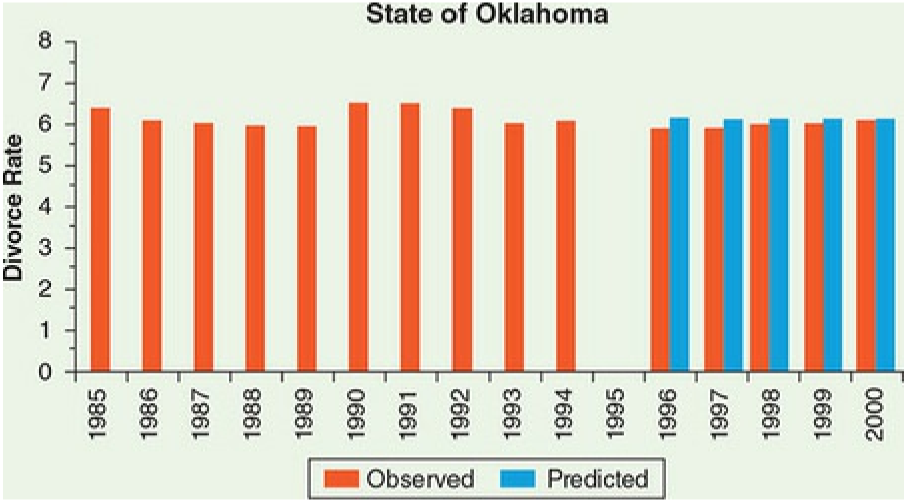
The most powerful type of quasi-experimental design that can be considered a before-and-after design is the regression–discontinuity design. This type of design can be used if participants are assigned to treatment solely based on a cutoff score on some assignment variable. Students may be admitted to a special intensive course based on a test score, or persons may become eligible for a housing voucher based on their income. In these two situations, test score and personal income are assignment variables.

Researchers using a regression–discontinuity design plot the relationship (the regression line) between scores on the assignment variable and the outcome of interest for those who did not enter the treatment as well as for those who did enter the treatment. If there is a jump in the regression line at the cutoff score, it indicates an effect of the treatment. For example, Sarah Kuck Jalbert, William Rhodes, Christopher Flygare, and Michael Kane (2010) studied the effect of reduced caseload size and intensive supervision on probation outcomes, using a regression–discontinuity design. Probationers were assigned to this program if they exceeded a certain value on a recidivism risk score. Jalbert and her colleagues found that at the point for probationers with risk scores near the cutoff score for program participation, recidivism dropped by 25.5% after 6 months for those who were admitted to the program.

Regression–discontinuity design: A quasi-experimental design in which individuals are assigned to a treatment and a comparison group solely on the basis of a cutoff score on some assignment variable, and then treatment effects are identified by a discontinuity in the regression line that displays the relation between the outcome and the assignment variable at the cutoff score.

Assignment variable: The variable used to specify a cutoff score for eligibility for some treatment in a regression–discontinuity design.

Exhibit 7.9 Divorce Rates in Oklahoma Before and After the Oklahoma City Bombing



Source: Nakonezny et al. 2004. “Did Divorces Decline After the Oklahoma City Bombing?” Journal of Marriage and Family, 66:90–100. Copyright © 2004, John Wiley & Sons. Reprinted with permission.

Summary: Causality in Quasi-Experiments

Let’s now examine how well quasi-experiments meet the criteria for identifying a nomothetic cause and the two additional challenges (introduced in Chapter 6):

Association between the hypothesized independent and dependent variables.

Quasi-experiments can provide evidence of association between the independent and dependent variables that is as unambiguous as that provided by a true experiment.

Time order of effects of one variable on the others.

This is a strength of the various quasi-experimental before-and-after designs, but we cannot be as sure of correctly identifying the time order of effects with nonequivalent control group designs because we cannot be certain that some features of the groups did not attract individuals to them who differed at the outset. This is a much greater problem with ex post facto control group designs.

Nonspurious relationships between variables.

We cannot entirely meet this challenge with a quasi-experimental design because we cannot be certain to rule out all potentially extraneous influences with either nonequivalent control group designs or before-and-after designs. Nonetheless, the criteria for these designs do give us considerable confidence that most extraneous influences could not have occurred. Ex post facto control group designs give us much less confidence about the occurrence of extraneous influences because of the likelihood of self-selection into the groups. This is why most researchers do not consider ex post facto control group designs to be quasi-experimental.

Mechanism that creates the causal effect.

The features of quasi-experiments and ex post facto designs do not in themselves allow identification of causal mechanisms; however, the repeated measures design does provide a means for testing hypotheses about the causal mechanism.

Context in which change occurs.

The quasi-experimental designs that involve multiple groups can provide a great deal of information about the importance of context, as long as the researcher measures contextual variables.

Validity in Experiments

Like any research design, experimental designs must be evaluated for their ability to yield valid conclusions. True experiments are particularly well suited for producing valid conclusions about causality (internal validity), but they are likely to fare less well in achieving generalizability. Quasi-experiments may provide more generalizable results than true experiments do, but they are more prone to problems of internal invalidity (although some quasi-experimental designs allow the researcher to rule out almost as many potential sources of internal invalidity as does a true experiment). It is important to distinguish nonequivalent control group designs from ex post facto designs when evaluating internal validity, given the problem of self-selection in ex post facto designs. Measurement validity is also a central concern, but experimental design does not in itself offer any special tools or particular advantages or disadvantages in measurement. In this section, you will learn more about the ways in which experiments help (or don't help) resolve potential problems of internal validity and generalizability (Campbell and Stanley 1966).

Causal (Internal) Validity

An experiment's ability to yield valid conclusions about causal effects is determined by the comparability of its experimental and comparison groups. First, of course, a comparison group must be created. Second, this comparison group must be so similar to the experimental group or groups that it can show in the posttest what the experimental group would have been like if it had not received the experimental treatment—if the independent variable had not varied. You now know that randomization is used to create a comparison group that is identical to the experimental group at the start of the experiment—with a certain margin of error that occurs with a process of random assignment. For this reason, a true experiment—a design with random assignment—is prone to fewer sources of internal invalidity than a quasi-experiment is.

Several sources of internal invalidity are considerably reduced by a research design that has a comparison group, but others are likely to occur unless a true experimental design with random assignment is used.

Sources of Internal Invalidity Reduced by a Comparison Group

The types of problem that can largely be eliminated by having a comparison group as well as a treatment group are those that arise during the study period itself (Campbell and Stanley 1966:8). Something unanticipated may happen (an effect of "history"), the pretest may have an unanticipated effect on subsequent posttests ("testing"), or the measurement instrument itself may perform differently a second time ("instrumentation"). Also, the participants themselves may change over time ("maturation") or simply decline to their normal levels of performance ("regression effects"). Each of these potential sources of internal invalidity is explained in more detail:

1. *History:* External events during the experiment (things that happen outside the experiment) can change subjects' outcome scores. Examples are newsworthy events that have to do with the focus of an experiment and major disasters to which subjects are exposed. This problem is often referred to as a history effect—that is, history during the experiment.

   Features of how the treatment is delivered can result in an effect of history, or external events, when there is a comparison, including in true and quasi-experimental designs. For example, in an experiment in which subjects go to a special location for a treatment, something in that location unrelated to the treatment might influence these subjects. Experimental and comparison group subjects in Richard Price, Michelle Van Ryn, and Amiram Vinokur's (1992) study of job search services differed in whether they attended the special seminars, so external events could have happened to subjects in the experimental group that might not have happened to those in the control group. Perhaps program participants witnessed a robbery outside the seminar building one day, and their orientations changed as a result. External events are a major concern in evaluation studies that compare programs in different cities or states (Hunt 1985:276–277).

2. *Testing:* Taking the pretest can in itself influence posttest scores. Subjects may learn something or be sensitized to an issue by the pretest and, as a result, respond differently the next time they are asked the same questions on the posttest.
3. *Instrumentation:* If the instrument used to measure the dependent variable changes in performance between the pretest and posttest, the result is termed a problem of instrumentation. For example, observers rating the behavior of students in a classroom may grow accustomed to the level of disruptions and therefore rate the same behavior as more appropriate in the posttest than they did in the pretest. This is similar to the problem of testing, except that it may be possible to reduce the effect of instrumentation by increasing control over the measurement instrument.
4. *Maturation:* Changes in outcome scores during experiments that involve a lengthy treatment period may result from maturation. Subjects may age, gain experience, or grow in knowledge, all as part of a natural maturational experience, and thus respond differently on the posttest than on the pretest.
5. *Regression:* People experience cyclical or episodic changes that result in different posttest scores, a phenomenon known as a regression effect. Subjects who are chosen for a study because they received very low scores on a test may show improvement in the posttest, on average, simply because some of the low scorers were having a bad day. Conversely, individuals selected for an experiment because they are suffering from tooth decay will not show improvement in the posttest because a decaying tooth is not likely to improve in the natural course of things. It is hard, in many cases, to know whether a phenomenon is subject to naturally occurring fluctuations, so the possibility of regression effects should be considered whenever subjects are selected because of their extremely high or low values on the outcome variable (Mohr 1992:56, 71–79).

History effect: A source of causal invalidity that occurs when events external to the study influence posttest scores; also called an effect of external events.

Regression effect: A source of causal invalidity that occurs when subjects who are chosen for a study because of their extreme scores on the dependent variable become less extreme on the posttest because of natural cyclical or episodic change in the variable.

History, testing, maturation, instrumentation, and regression effects could explain any change over time in most of the before-and-after designs because these designs do not have a comparison group. Repeated measures panel studies and time series designs are better in this regard because they allow the researcher to trace the pattern of change or stability in the dependent variable until and after the treatment. However, it is much more desirable to have a comparison group that, like the treatment group, will also be affected by these sources of internal invalidity and so can control for them. Of course, these factors are not a problem in true experiments because those have an experimental group and the comparison group.

Sources of Internal Invalidity Reduced by Randomization

You have already learned that the purpose of randomization, or random assignment to the experimental and comparison groups, is to equate the two or more groups at the start of the experiment. The goal is to eliminate the effect of selection bias.

Selection bias: A source of internal (causal) invalidity that occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.

> 6. *Selection bias:* The composition of the experimental and comparison groups in a true experiment is unlikely to be affected by selection bias. Randomization equates the groups' characteristics, although with some possibility for error due to chance. The likelihood of difference due to chance can be identified with appropriate statistics.
>
> When subjects are not assigned randomly to treatment and comparison groups, as in nonequivalent control group designs, the threat of selection bias is very great. Even if the researcher selects a comparison group that matches the treatment group on important variables, there is no guarantee that the groups were similar initially in the dependent variable or in some other characteristic that ultimately influences posttest scores. However, a pretest helps the researchers determine and control for selection bias. Because most variables that might influence outcome scores will also have influenced scores on the pretest, statistically controlling for the pretest scores also controls many of the unmeasured variables that might have influenced the posttest scores. The potential for selection bias is much greater with an ex post facto control group design because participants have the ability to select the group they enter based on the treatment they expect to receive.

Sources of Internal Invalidity That Require Attention While the Experiment Is in Progress

Even in a research design that involves a comparison group and random assignment, whether or not there is a pretest, the experimental and comparison groups can become different over time because of changes in group membership, interaction between members of the experimental and comparison groups, or effects that are related to the treatment but are not the treatment itself.

7. *Differential attrition:* This problem occurs when the groups become different after the experiment begins because more participants drop out of one of the groups than out of the other(s) for various reasons. Differential attrition (mortality) is not a likely problem in a laboratory experiment that occurs in one session, such as Czopp et al.'s (2006) experiment with college students, but some experiments continue over time. Subjects who experience the experimental condition may become more motivated than comparison subjects are to continue in the experiment and so be less likely to drop out.

You learned in Chapter 6 that attrition can be a major problem in longitudinal designs, whether these are simple panel studies or quasi-experimental repeated measures designs. When many subjects have left a panel between the pretest and posttest (or between any repeated measures in the study), a comparison of the average differences between subjects at the start and end of the study period may mislead us to think that the subjects have changed, when what actually happened is that subjects who had dropped out of the study were different from those who remained in it. For example, people with less education and who are not married have been less likely to continue in the large ongoing Panel Study of Income Dynamics (Lillard and Panis 1998:442). Statistical adjustments can reduce the effects of panel attrition, but it is always important to compare the characteristics of study dropouts and those who remain for follow-up.

When the independent variable in an experimental study is a treatment or an exposure to something over time, bias caused by differential attrition can be reduced by using intent-to-treat analysis. In this type of analysis, outcomes are compared between all participants who started in the experimental group—those who were intended to receive the treatment—and all participants who started in the control group, whether or not any of these participants left the study before the treatment was fully delivered (Shrout 2011:7). Of course, an intent-to-treat analysis is likely to reduce the researcher's estimate of the treatment effect because some participants will not have received the full treatment, but it provides a more realistic estimate of effects that are likely to occur if the treatment is administered to persons who can leave before the experiment is over (Bloom 2008:120).

8. *Contamination:* When the comparison group in an experiment is in some way affected by, or affects, the treatment group, there is a problem with contamination. Contamination is not ruled out by the basic features of experimental and quasi-experimental designs, but careful inspection of the research design can determine

how much it is likely to be a problem in a particular experiment. This problem basically arises from the failure to control adequately the conditions of the experiment. If the experiment is conducted in a laboratory, if members of the experimental group and the comparison group have no contact while the study is in progress, and if the treatment is relatively brief, contamination is not likely to be a problem. To the degree that these conditions are not met, the likelihood of contamination will increase. For example, contamination was a potential problem in a field-based study by Price et al. (1992) about the effects of a job search training program on the risk of depression among newly unemployed persons. Because the members of both the experimental group (who received the training) and the control group (who did not receive the training) used the same unemployment offices, they could have talked to each other about their experiences while the study was in progress.

9. *Compensatory rivalry:* A problem related to contamination, also termed the John Henry effect, can occur when comparison group members are aware that they are being denied some advantages and, in response, increase their efforts to compensate for this denial (Cook and Campbell 1979:55).

10. *Demoralization:* This problem involves the opposite reaction to compensatory rivalry; that is, comparison group participants discover that they are being denied some treatments they believe are valuable and as a result they feel demoralized and perform worse than expected. Both compensatory rivalry and demoralization thus distort the impact of the experimental treatment.

11. *Expectancies of experimental staff:* Change among experimental subjects may result from the positive expectations of the experimental staff who are delivering the treatment rather than from the treatment itself. Expectancies of experimental staff may alter the experimental results if staff—even well-trained staff—convey their enthusiasm for an experimental program to the subjects in subtle ways. This is a special concern in evaluation research, when program staff and researchers may be biased in favor of the program for which they work and eager to believe that their work is helping clients. Such positive staff expectations thus create a self-fulfilling prophecy. However, in experiments on the effects of treatments such as medical drugs, double-blind procedures can be used: Staff delivering the treatments do not know which subjects are getting the treatment and which are receiving a placebo, something that looks like the treatment but has no effect.

12. *Placebo effect:* Treatment misidentification may occur when subjects receive a treatment that they consider likely to be beneficial and improve because of that expectation rather than because of the treatment itself. In medical research, where the placebo effect often results from a chemically inert substance that looks like the experimental drug but actually has no direct physiological effect, some research has indicated that the placebo effect itself produces positive health effects in many patients suffering from relatively mild medical problems (Goleman 1993a:C3). It is not clear that these improvements are really any greater than what the patients would

have experienced without the placebo (Hrobjartsson and Gotzsche 2001). In any case, it is possible for placebo effects to occur in social science research also, so, when possible, experimental researchers can reduce this threat to internal validity by treating the comparison group with something that seems similar to what the experimental group receives. You read earlier about the short feedback that Czopp et al. (2006:795) had their "control" subjects receive to give them an experience similar to the "confrontation" feedback that the experimental subjects received.

13. *Hawthorne effect:* Members of the treatment group may change relative to the dependent variable because their participation in the study makes them feel special. This problem can occur when treatment group members compare their situation with that of members of the control group, who are not receiving the treatment; in this situation, it would be a type of contamination effect. But experimental group members might feel special simply because they are in the experiment. This is termed a Hawthorne effect, after a famous productivity experiment at the Hawthorne electric plant outside Chicago. As the story has been told, the workers worked harder no matter what physical or economic conditions the researchers changed to influence productivity; the motivation for the harder work simply seemed to be that the workers felt special because of being in the experiment (Whyte 1955:34).

Let me quickly add that a careful review of the actual Hawthorne results shows that there wasn't really a clear effect of participating in that experiment (Jones 1992). The Hawthorne effect was itself mostly a matter of misinterpretation and hype. But we can never ignore the possibility that participation in an experiment may itself change participants' orientations and behavior. This is a particular concern in evaluation research when program clients know that the research findings may affect the chances for further program funding.

Differential attrition (mortality): A problem that occurs in experiments when comparison groups become different because subjects are more likely to drop out of one of the groups for various reasons.

Intent-to-treat analysis: When analysis of the effect of a treatment on outcomes in an experimental design compares outcomes for all those who were assigned to the treatment group with outcomes for all those who were assigned to the control group, whether or not participants remained in the treatment group.

Contamination: A source of causal invalidity that occurs when either the experimental or the comparison group is aware of the other group and is influenced in the posttest as a result.

Compensatory rivalry (John Henry effect): A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied some advantages and increase their efforts by way of compensation.

Demoralization: A type of contamination in experimental and quasi-experimental designs that occurs when control group members feel they have been left out of some valuable treatment and perform worse as a result.

Expectancies of experimental staff: A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects results from the positive expectancies of the staff who are delivering the treatment rather than from the treatment itself; also called a *self-fulfilling prophecy.*

Double-blind procedure: An experimental method in which neither the subjects nor the staff delivering experimental treatments know which subjects are getting the treatment and which are receiving a placebo.

Placebo effect: A source of treatment misidentification that can occur when subjects receive a fake "treatment" they think is beneficial and improve because of that expectation even though they did not receive the actual treatment or received a treatment that had no real effect.

Hawthorne effect: A type of contamination in research designs that occurs when members of the treatment group change relative to the dependent variable because their participation in the study makes them feel special.

Generalizability

The need for generalizable findings can be thought of as the Achilles heel of true experimental design. The design components that are essential for a true experiment and that minimize the threats to causal validity make it more difficult to achieve sample generalizability (being able to apply the findings to some clearly defined larger population) and cross-population generalizability (generalizing across subgroups and to other populations and settings). As a result, findings are often not replicated when experiments are repeated in different settings (Finkel et al. 2017)—as you learned in relation to the Sherman and Berk experiments on the police response to domestic violence (see Chapter 2). Nonetheless, no one conducts experiments just to find out how freshman psychology students react to confrontation (or some other experimental "treatment") at your university. Experimental researchers are seeking to learn about general processes, so we have to consider ways to improve the generalizability of their results.

Sample Generalizability

Subjects who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the study's duration are unlikely to be a representative sample of any large population of interest to social scientists. Can they be expected to react to the experimental treatment in the same way as members of the larger population? The generalizability of the treatment and of the setting for the experiment also must be considered (Cook and Campbell 1979:73–74). The more artificial the experimental arrangements are, the greater the problem will be (Campbell and Stanley 1966:20–21).

A researcher can take steps both before and after an experiment to increase a study's generalizability. Conducting a field experiment, such as Sherman and Berk's (1984) study of arrest in actual domestic violence incidents, is likely to yield more generalizable findings than are laboratory experiments, for which subjects must volunteer. In some field experiments, participants can even be selected randomly from the population of interest, and, thus, the researchers can achieve results generalizable to that population. For example, some studies of the effects of income supports on the work behavior of poor persons have randomly sampled persons within particular states before randomly assigning them to experimental and comparison groups. When random selection is not feasible, the researchers may be able to increase generalizability by selecting several different experimental sites that offer marked contrasts on key variables (Cook and Campbell 1979:76–77).

Conducting research online is another way to involve people other than undergraduate students in research studies, but bear in mind that Internet users are more likely to be college educated and male than the general population (Hewson, Vogel, and Laurent 2016:79). WEXTOR (http://wextor.org/wextor/en/) provides an online tool for generating experiments both online and in the laboratory (Reips and Neuhaus 2002), while Mechanical Turk provides easy access to people who have signed up to conduct tasks online and can be recruited for online experiments (https://www.mturk.com/mturk/welcome). Extensive pretesting of web-based experiments is critical to identify and resolve problems (Hewson et al. 2016:136137).

Factorial Surveys

Factorial surveys embed the features of true experiments into a survey design to maximize generalizability. In the most common type of factorial survey, respondents are asked for their likely responses to one or more vignettes about hypothetical situations. The content of these vignettes is varied randomly among survey respondents to create "treatment groups" that differ in particular variables reflected in the vignettes.

Factorial survey: A survey in which randomly selected subsets of respondents are asked different questions, or are asked to respond to different vignettes, to determine the causal effect of the variables represented by these differences.

Greet Van Hoye and Filip Lievens (2003) used a factorial survey design to test the effect of job applicants' sexual orientation on ratings of their hirability by professionals who make personnel decisions. Van Hoye and Lievens first identified 252 actual selection professionals—people involved daily in personnel selection and recruitment—from consulting firms and company human resource departments. The researchers mailed to each of these professionals a packet with four items: (1) a *letter* inviting their participation in the study, (2) a *job posting* that described a company and a particular job opening in that company, (3) a *candidate profile* that described someone ostensibly seeking that job, and (4) a *response form* on which the selection professionals could rate the candidate's hirability.

The experimental component of the survey was created by varying the candidate profiles. Van Hoye and Lievens created nine different candidate profiles. Each profile used very similar language to describe a candidate's gender (they were all male), age, nationality, family situation, education, professional experience, and personality. However, the family situations were varied to distinguish candidates as heterosexual, homosexual, or "possibly homosexual"—single and older than 30. Other characteristics were varied to distinguish candidates who were "poor," "moderate," and "excellent" matches to the job opening. An example of a profile for a homosexual male who was a "good" candidate for the job included the following language:

1. Personal Data

> Name: Peter Verschaeve
> Gender: Male
> Age: 33 years
> Family situation: Living together with John Vermeulen, fashion designer
>
> 2. Educational and Professional Experience
>   1990–1993: PUC Diepenbeek—MBA, Marketing Major
>   1991–now: Human resources manager of an electronics manufacturer
> 3. Personality
>   Peter Verschaeve is self-assured and assertive. He interacts with others in a friendly and warm manner. (Van Hoye and Lievens 2003:27)

The combination of three different descriptions of family situation and three different levels of candidate quality resulted in nine different candidate profiles. Each selection professional was randomly assigned to receive one of these nine candidate profiles. As a result, there was no relationship between who a particular selection professional was and the type of candidate profile he or she received.

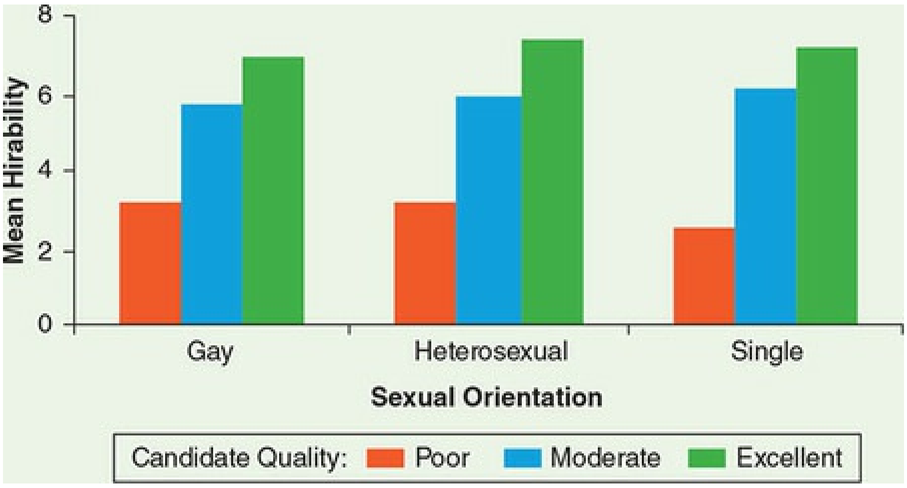
The results of the study appear in Exhibit 7.10. The average hirability ratings did not differ between candidates who were gay, heterosexual, or single, but hirability increased in direct relation to candidate quality. Van Hoye and Lievens (2003:26) concluded that selection professionals based their evaluations of written candidate profiles on candidate quality, not on their sexual orientation—at least in Flanders, Belgium.

Because Van Hoye and Lievens surveyed real selection professionals at their actual workplaces, we can feel more comfortable with the generalizability of their results than if they had just recruited college students for an experiment in a laboratory. In a second experiment, Pedulla (2016) also used a factorial survey design to test the reactions of hiring managers to candidates with different employment histories and ask about the reasons behind their reactions. However, there is still an important limitation to the generalizability of factorial surveys such as these: A factorial survey research design indicates only what respondents say they would do in situations that have been described to them. If the selection professionals had to make a recommendation for hiring to an actual employer, we cannot be sure that they would act in the same way. So factorial surveys do not completely resolve the problems caused by the difficulty of conducting true experiments with representative samples. Nonetheless, by combining some of the advantages of experimental and survey designs, factorial surveys can provide stronger tests of causal hypotheses than can other surveys and more generalizable findings than can experiments.

External Validity

Researchers are often interested in determining whether treatment effects identified in an experiment hold true for subgroups of subjects and across different populations, times, or settings. Of course, determining that a relationship between the treatment and the outcome variable holds true for certain subgroups does not establish that the relationship also holds true for these subgroups in the larger population, but it suggests that the relationship might be externally valid.

Exhibit 7.10 Average Hirability Ratings in Relation to Candidate Quality and Sexual Orientation



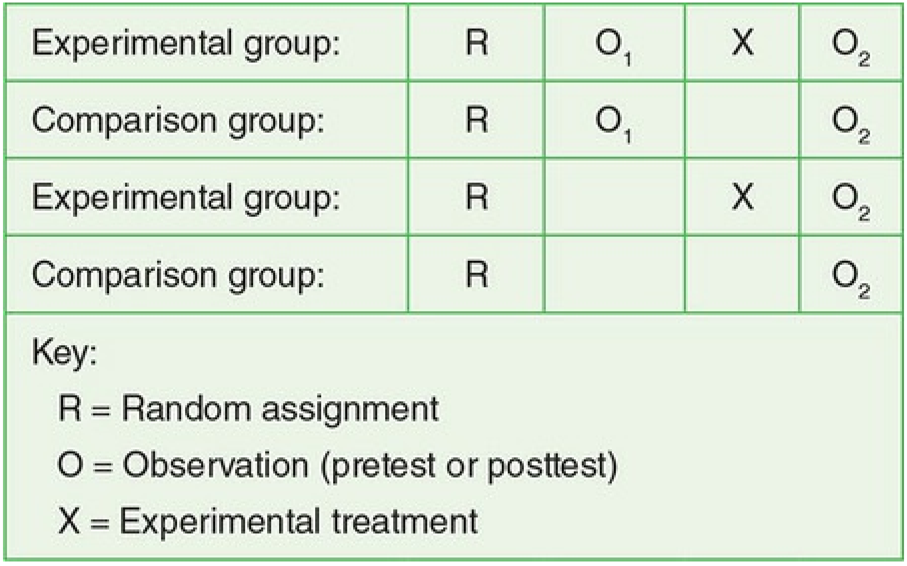
Source: Van Hoye, Greet, and Filip Lievens. 2003. "The Effects of Sexual Orientation on Hirability Ratings: An Experimental Study." Journal of Business and Psychology 18:15–30. Copyright © 2003, Human Science Press, Inc. Reprinted with permission from Springer.

We have already seen examples of how the existence of treatment effects in particular subgroups of experimental subjects can help us predict the cross-population generalizability of the findings. For example, Sherman and Berk's (1984) research (see Chapter 2) found that arrest did not deter subsequent domestic violence for unemployed individuals; arrest also failed to deter subsequent violence in communities with high levels of unemployment. Price et al. (1992) found that intensive job search assistance reduced depression among individuals who were at high risk for it because of other psychosocial characteristics; however, the intervention did not influence the rate of depression among individuals at low risk for depression. This is an important interaction effect that limits the generalizability of the treatment, even if Price et al.'s sample was representative of the population of unemployed persons.

Interaction of Testing and Treatment

A variant on the problem of external validity occurs when the experimental treatment has an effect only when particular conditions created by the experiment occur. One such problem occurs when the treatment has an effect only if subjects have had the pretest. The pretest sensitizes the subjects to some issue, so that when they are exposed to the treatment, they react in a way that differs from how they would have reacted had they not taken the pretest. In other words, testing and treatment interact to produce the outcome. For example, answering questions in a pretest about racial prejudice may sensitize subjects so that when exposed to the experimental treatment, seeing a film about prejudice, their attitudes are different from what they would have been otherwise. In this situation, the treatment truly had an effect, but it would not have had an effect had it been provided without the sensitizing pretest. This possibility can be evaluated with the Solomon four-group design described earlier (see Exhibit 7.11).

Exhibit 7.11 Solomon Four-Group Design Testing the Interaction of Pretesting and Treatment



Experimental group:	R	O_1	X	O_2
Comparison group:	R	O_1		O_2
Experimental group:	R		X	O_2
Comparison group:	R			O_2

Key:
R = Random assignment
O = Observation (pretest or posttest)
X = Experimental treatment

As you can see, no single procedure establishes the external validity of experimental results. Ultimately, we must base our evaluation of external validity on the success of replications taking place at different times and places and using different forms of the treatment.

There is always an implicit trade-off in experimental design between maximizing causal validity and generalizability. The more the assignment to treatments is randomized and all experimental conditions are controlled, the less likely it is that the research subjects and setting will be representative of the larger population. College students are easy to recruit and assign to artificial but controlled manipulations, but both practical and ethical concerns preclude this approach with many groups and with respect to many treatments. However, although we need to be skeptical about the generalizability of the results of a single experimental test of a hypothesis, the body of findings accumulated from many experimental tests with different people in different settings can provide a solid basis for generalization (Campbell and Russo 1999:143).

Ethical Issues in Experimental Research

Social science experiments can raise difficult ethical issues. You have already read in Chapter 3 that Philip Zimbardo (2004:34) ended his Stanford prison experiment after only 6 days, rather than after the planned 2 weeks, because of the psychological harm that seemed to result from the unexpectedly sadistic behavior of some of the "guards." Although Zimbardo's follow-up research convinced him that there had been no lasting harm to subjects, concern about the potential for harm would preclude many such experiments today.

Nonetheless, experimental research continues because of the need for very good evidence about cause–effect relationships to inform social theory as well as social policy. The particular strength of randomized experiments for answering causal questions means that they can potentially prevent confusion in social theory and avoid wasting time and resources on ineffective social programs (Mark and Gamble 2009:203).

Two ethical issues are of special importance in experimental research designs. Deception is an essential part of many experimental designs, despite the ethical standard of subjects' informed consents. As a result, contentious debate continues about the interpretation of this standard. In addition, experimental evaluations of social programs pose ethical dilemmas because they require researchers to withhold possibly beneficial treatment from some of the subjects just on the basis of chance (Boruch 1997). In this section, I give special attention to the problems of deception and the distribution of benefits in experimental research.

Deception

Deception is used in social experiments to create more "realistic" treatments, often within the confines of a laboratory. Chapter 3 described Milgram's (1965) use of deception in his classic study of obedience to authority. Volunteers were recruited for what they were told was a study of the learning process, not a study of "obedience to authority." The experimenter told the volunteers that they were administering electric shocks to a "student" in the next room, when there were actually neither students nor shocks. Most subjects seemed to believe the deception.

You learned in Chapter 3 that Milgram's description of dehoaxing inflated the consistency with which it was used and the amount of deception that was revealed. You also learned that the dehoaxing increased the negative impact of the experimental experience for some participants. You may therefore be reassured to know that experiments by Davide Barrera and Brent Simpson (2012) found that students' experience of being deceived in a social psychology experiment did not affect their subsequent behavior toward others.

Whether or not you believe that you could be deceived in this way, you are not likely to be invited to participate in an experiment such as Milgram's. Current federal regulations preclude deception in research that might trigger such upsetting feelings. However, deception such as that used by Czopp and his colleagues (2006) is still routine in social psychology laboratories. Deceiving students that they were working with another student whom they could not see was essential to their manipulation of interpersonal confrontation. In this experiment, as in many others, the results would be worthless if subjects understood what was really happening to them. The real question is "Is this sufficient justification to allow the use of deception?"

The American Sociological Association's (ASA) *Code of Ethics and Policies and Procedures of the ASA Committee on Professional Ethics* (1999) does not discuss experimentation explicitly, but it does highlight the ethical dilemma posed by deceptive research:

> 12.05 Use of Deception in Research: (a) Sociologists do not use deceptive techniques (1) unless they have determined that their use will not be harmful to research participants; is justified by the study's prospective scientific, educational, or applied value; and that equally effective alternative procedures that do not use deception are not feasible, and (2) unless they have obtained the approval of institutional review boards or, in the absence of such boards, with another authoritative body with expertise on the ethics of research. (b) Sociologists never deceive research participants about significant aspects of the research that would affect their willingness to participate, such as physical risks, discomfort, or unpleasant emotional experiences. (c) When deception is an integral feature of

> the design and conduct of research, sociologists attempt to correct any misconception that research participants may have no longer than at the conclusion of the research. (p. 16)

Thus, the ASA approach is to allow deception when it is unlikely to cause harm, is necessary for the research, and is followed by adequate explanation after the experiment is over. As you learned in Chapter 3, the newly revised federal human subjects research standards make an explicit exemption for "research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses . . . or audiovisual recording" when the subject had agreed to the intervention (*Federal Register* 2017:7261, 7262, 7264). This would exempt the typical social psychology laboratory experiment from review, but not experiments involving interventions like Milgram's or Zimbardo's that would not be considered "benign."

David Willer and Henry A. Walker (2007:"Debriefing") pay particular attention to debriefing after deception in their book about experimental research. They argue that every experiment involving deception should be followed immediately for each participant with dehoaxing, in which the deception is explained, and then by desensitization, in which all the participants' questions are answered to their satisfaction and those participants who still feel aggrieved are directed to a university authority to file a complaint or to a counselor for help with their feelings. This is sound advice.

Debriefing is a special concern in experiments conducted on the web, since participants who withdraw early are usually no longer available. It is therefore good practice to require participants in web experiments (and surveys) to provide an e-mail address initially so that they can be contacted again. The website where the experiment is provided should also be programmed so that termination before the experiment ends takes the participant immediately to a debriefing site, where information about the experiment is provided (Toepoel 2016:46).

Selective Distribution of Benefits

Field experiments conducted to evaluate social programs also can involve issues of informed consent (Hunt 1985:275–276). One ethical issue that is somewhat unique to field experiments is the selective distribution of benefits: How much are subjects harmed by the way treatments are distributed in the experiment? For example, Sherman and Berk's (1984) experiment, and its successors, required police to make arrests in domestic violence cases largely on the basis of a random process. When arrests were not made, did the subjects' abused spouses suffer? Price et al. (1992) randomly assigned unemployed individuals who had volunteered for job search help to an intensive program. Were the unemployed volunteers assigned to the comparison group at a big disadvantage?

Selective distribution of benefits: An ethical issue about how much researchers can influence the benefits subjects receive as part of the treatment being studied in a field experiment.

Is it ethical to give some potentially advantageous or disadvantageous treatment to people on a random basis? Random distribution of benefits is justified when the researchers do not know whether some treatment actually is beneficial or not—and, of course, it is the goal of the experiment to find out (Mark and Gamble 2009:205). Chance is as reasonable a basis for distributing the treatment as any other. Also, if insufficient resources are available to fully fund a benefit for every eligible person, then distribution of the benefit on the basis of chance to equally needy persons is ethically defensible (Boruch 1997:66–67).

The extent to which participation was voluntary varied in the field studies discussed in this chapter. Potential participants in the Price et al. (1992) study of job search training for unemployed persons signed a detailed consent form in which they agreed to participate in a study involving random assignment to one of the two types of job search help. However, researchers only accepted into the study persons who expressed equal preference for the job search seminar and the mailed job materials used for the control group.

Thus, Price et al. (1992) avoided the problem of not acceding to subjects' preferences. It, therefore, doesn't seem at all unethical that the researchers gave treatment to only some of the subjects. As it turned out, subjects did benefit from the experimental treatment (the workshops). Now that the study has been conducted, government bodies will have a basis for expecting that tax dollars spent on job search workshops for the unemployed will have a beneficial impact. If this knowledge results in more such programs, the benefit of the experiment will have been considerable, indeed.

Unlike the subjects in the Price et al. (1992) study, individuals who were the subjects of domestic violence complaints in the Sherman and Berk (1984) study had no choice about being arrested or receiving a warning, nor were they aware that they were in a research study.

Perhaps it seems unreasonable to let a random procedure determine how police resolve cases of domestic violence. And, indeed, it would be unreasonable if this procedure were a regular police practice. The Sherman and Berk (1984) experiment and its successors do pass ethical muster, however, when seen for what they were: a way of learning how to increase the effectiveness of police responses to this all-too-common crime (Mark and Gamble 2009:205). The initial Sherman and Berk findings encouraged police departments to make many more arrests for these crimes, and the follow-up studies resulted in a better understanding of when arrests are not likely to be effective. The implications of this research may be complex and difficult to implement, but the research provides a much stronger factual basis for policy development.

Conclusions

True experiments play two critical roles in social science research. First, they are the best research design for testing nomothetic causal hypotheses. Even when conditions preclude use of a true experimental design, many research designs can be improved by adding some experimental components. Second, true experiments also provide a comparison point for evaluating the ability of other research designs to achieve causally valid results.

Despite obvious strengths, true experiments are used infrequently to study many research problems that interest social scientists. There are three basic reasons: (1) The experiments required to test many important hypotheses require far more resources than are often available; (2) many research problems of interest to social scientists are not amenable to experimental designs, for reasons ranging from ethical considerations to the limited possibilities for randomly assigning people to different conditions in the real world; and (3) the requirements of experimental design usually preclude large-scale studies and so limit generalizability to a degree that is unacceptable to many social scientists. Quasi-experiments can be an excellent design alternative.

Although it may be possible to test a hypothesis with an experiment, it may not always be desirable to do so. When a social program is first being developed and its elements are in flux, it is not a good idea to begin a large evaluation study that cannot possibly succeed unless the program design remains constant. Researchers should wait until the program design stabilizes somewhat. It also does not make sense for evaluation researchers to test the impact of programs that cannot actually be implemented or that are unlikely to be implemented in the real world because of financial or political problems (Rossi and Freeman 1989:304–307).

Even laboratory experiments are inadvisable when they do not test the real hypothesis of interest, but test instead a limited version amenable to laboratory manipulation. The intersecting complexity of societies, social relationships, and social beings—of people and the groups to which they belong—is so great that it often defies reduction to the simplicity of a laboratory or restriction to the requirements of experimental design. Yet the virtues of experimental designs mean that they should always be considered when explanatory research is planned.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Assignment variable 236
Before-and-after design 231
Comparison group 224
Compensatory rivalry (John Henry effect) 241
Contamination 241
Control group 224
Demoralization 241
Differential attrition (mortality) 240
Double-blind procedure 241
Ex post facto control group design 231
Expectancies of experimental staff 241
Experimental group 224
External events 238
Factorial survey 243
Field experiment 230
Hawthorne effect 242
History effect 238
Intent-to-treat analysis 240
Matching 227
Multiple group before-and-after design 235
Nonequivalent control group design 231
Placebo effect 241
Posttest 225
Pretest 225
Quasi-experimental design 231
Randomized comparative change design 225
Randomized comparative posttest design 226
Regression–discontinuity design 236
Regression effect 239
Repeated measures panel design 235
Selection bias 239
Selective distribution of benefits 247
Solomon four-group design 228
Time series design 235
True experiment 224

Highlights

- The independent variable in an experiment is represented by a treatment or other intervention.

Some subjects receive one type of treatment; others may receive a different treatment or no treatment. In true experiments, subjects are assigned randomly to comparison groups.

- Experimental research designs have three essential components: (1) use of at least two groups of subjects for comparison, (2) measurement of the change that occurs as a result of the experimental treatment, and (3) use of random assignment. In addition, experiments may include identification of a causal mechanism and control over experimental conditions.
- Random assignment of subjects to experimental and comparison groups eliminates systematic bias in group assignment. The odds of a difference between the experimental and comparison groups because of chance can be calculated. These chances become very small for experiments with at least 30 subjects per group.
- Both random assignment and random sampling rely on a chance selection procedure, but their purposes differ. Random assignment involves placing predesignated subjects into two or more groups on the basis of chance; random sampling involves selecting subjects out of a larger population on the basis of chance. Matching of cases in the experimental and comparison groups is a poor substitute for randomization because identifying in advance all important variables on which to make the match is not possible. However, matching can improve the comparability of groups, when it is used to supplement randomization.
- Quasi-experiments include features that maximize the comparability of the control and experimental groups and make it unlikely that self-selection determines group membership.
- Causal conclusions derived from experiments can be invalid because of influences including selection bias, effects of external events, and cross-group contamination. In true experiments, randomization, use of a comparison group, and pretests and posttests should eliminate most of these sources of internal invalidity. However, when conditions are not carefully controlled during an experiment, differential attrition, contamination, compensatory rivalry, and demoralization can create differences between groups and threaten the validity of causal conclusions.
- Quasi-experiments may provide more generalizable results than true experiments do, but they are more prone to some problems of internal invalidity because of their lack of random assignment (although some quasi-experimental designs allow the researcher to rule out almost as many potential sources of internal invalidity as does a true experiment). The generalizability of experimental results declines if the study conditions are artificial and the experimental subjects are unique. Field experiments are likely to produce more generalizable results than experiments conducted in the laboratory.
- The external validity of causal conclusions is determined by the extent to which they apply to different types of individuals and settings. When causal conclusions do not apply to all the subgroups in a study, they are not generalizable to corresponding subgroups in the population—and so they are not externally valid with respect to those subgroups. Causal conclusions can also be considered externally invalid when they occur only under the experimental conditions.
- Subject deception is common in laboratory experiments and poses unique ethical issues. Researchers must weigh the potential harm to subjects and debrief subjects who have been deceived. In field experiments, a common ethical problem is selective distribution of benefits. Random assignment may be the fairest way of allocating treatment when treatment openings are insufficient for all eligible individuals and when the efficacy of the treatment is unknown.

Discussion Questions

1. Review Pedulla’s (2016) “employment history” experiment with which this chapter began. Diagram the experiment using the exhibits in this chapter as a model (use the callback experiment, not the survey experiment). Discuss the extent to which experimental conditions were controlled and the causal mechanism was identified. How confident can you be in the causal conclusions from the study, based on review of the threats to internal validity discussed in this chapter? How generalizable do you think the study’s results are to the population from which cases were selected? To specific subgroups in the study? How thoroughly do the researchers discuss these issues?
2. Describe a true experiment that could investigate the effect of the “nonstandard, mismatched positions penalty” using a laboratory experiment, rather than Pedulla’s (2016) field experiment design. What would the advantages be to conducting an experiment like this in a laboratory setting (such as on your campus)? What problems do you envision in implementing such a design? What would make you worry about generalizability of your findings?
3. Do you think that the government should fund studies like Pedulla’s in order to identify biases and other barriers in the labor market? Are there some other employment issues that you think should be studied in this way? Explain your reasoning.

Practice Exercises

1. Arrange with an instructor in a large class to conduct a multiple pretest–posttest study of the impact of watching a regularly scheduled class movie. Design a 10-question questionnaire to measure knowledge about the topics in the film. Administer this questionnaire shortly before and shortly after the film is shown and then again 1 week afterward. After scoring the knowledge tests, describe the immediate and long-term impact of the movie.
2. Volunteer for an experiment! Contact the psychology department, and ask about opportunities for participating in laboratory experiments. Discuss the experience with your classmates.
3. Take a few minutes to review the “Sources of Internal Invalidity” lesson from the “Interactive Exercises” link on the study site. It will be time well spent.
4. Select an article that used an experimental design from the book’s study site, at edge.sagepub.com/schutt9e. Diagram the design and identify sources of internal and external invalidity that are not controlled by the experimental design.

Ethics Questions

1. What specific rules do you think should guide researchers' decisions about subject deception and the selective distribution of benefits? How much deception should be allowed, and under what circumstances? Was the deception in the Milgram study acceptable? What about deception in "Psych 101" lab experiments? Do you think it would be acceptable to distribute free bicycles to a random selection of commuters for 2 years to see if they change their habits more than do those in a matched control group? What about the complaints of a member of the control group (after all, she pays her taxes, too!)?
2. Under what conditions do you think that the randomized assignment of subjects to a specific treatment is ethical in social science research? Was it ethical for Sherman and Berk (1984) and the researchers who conducted the replication studies to randomly assign individuals accused of domestic violence to an arrest or nonarrest treatment? What about randomly assigning some welfare recipients to receive higher payments than others? And what about randomly assigning some students to receive a different instructional method than others?

Web Exercises

1. Go to Sociosite at www.sociosite.net/index.php. Choose "Subject Areas." Choose a sociological subject area you are interested in. How would you conduct a study on your chosen subject using experimental methods? Choose at least five of the key terms listed at the end of this chapter that are relevant to and incorporated in the research experiment you have located on the web. Explain how each of the five key terms you have chosen plays a role in the research example you have found on the web.
2. Try out the process of randomization. Go to the website www.randomizer.org. Type numbers into the randomizer for an experiment with 1 group ("sets") and 40 individuals ("numbers") per set, with a number range from 1 to 2. Specify that each number in a set is not to be unique ("No"), and the list should be sorted from least to greatest, and printed with "Place Markers Within." Now click "RANDOMIZE NOW." Repeat the process for an experiment with 1 group and 40 numbers per set (range is now 1 to 4). Does the distribution of numbers assigned to each group seem to be random?
3. Participate in a social psychology experiment on the web. Go to www.socialpsychology.org/expts.htm. Pick an experiment in which to participate and follow the instructions. After you finish, write up a description of the experiment and evaluate it using the criteria discussed in the chapter.

Video Interview Questions

Listen to the researcher interview for Chapter 7 at edge.sagepub.com/schutt9e.

1. Why was it important for the research assistant to use a script in this study?
2. How did Professor Youngreen measure creative output in his study?

SPSS Exercises

Because the GSS2016 doesn't provide experimental data to work with, we'll pause in our study of support for capital punishment and examine some relationships involving workplace variables such as some of those that were the focus of research reviewed in this chapter.

Do the features of work influence attitudes about the work experience? We can test some hypothetical answers to this question with the GSS2016 data set (although not within the context of an experimental design).

1. Describe the feelings of working Americans about their jobs and economic rewards, based on their responses to questions about balancing work and family demands, their satisfaction with their finances, and their job satisfaction. Generate the frequencies as follows:
    a. Click Analyze/Descriptive statistics/Frequencies.
    b. Select SATFIN, SATJOB.

       How satisfied are working people with their jobs and their pay?
    c. Do these feelings vary with work features?
        a. Pose at least three hypotheses in which either SATFIN or SATJOB is the dependent variable and one of the following two variables is the independent variable: earnings or work status. Now test these hypotheses by comparing average scores on the attitudinal variables between categories of the independent variables:
            i. Click Analyze/Compare Means/Means
            ii. Select Dependent List: SATFIN, SATJOB
            iii. Independent List: INCOMEFAM4, WRKSTAT
        b. Which hypotheses appear to be supported? (Remember to review the distributions of the dependent variables [SATFIN, SATJOB] to remind yourself what a higher average score indicates on each variable.)

Developing a Research Proposal

Your work in this section should build on your answers to the proposal development questions in the last chapter, assuming that you will use an experimental design (Exhibit 3.10, #13, #14, #17).

1. Design a laboratory experiment to test one of your hypotheses or a related hypothesis. Describe the experimental design, commenting on each component of a true experiment. Specify clearly how the independent variable will be manipulated and how the dependent variable will be measured.
2. Assume that your experiment will be conducted on campus. Formulate recruitment and randomization procedures.
3. Discuss the extent to which each source of internal invalidity is a problem in the study. Propose procedures to cope with these sources of invalidity.
4. How generalizable would you expect the study's findings to be? What can be done to increase generalizability?
5. Develop appropriate procedures for the protection of human subjects in your experiment. Include among these procedures a consent form. Give particular attention to any aspects of the study that are likely to raise ethical concerns.

Chapter 8 Survey Research

- Research That Matters, Questions That Count
- Survey Research in the Social Sciences
  - Attractions of Survey Research
    - *Versatility*
    - *Efficiency*
    - *Generalizability*
  - The Omnibus Survey
  - Errors in Survey Research
- Writing Survey Questions
  - Avoid Confusing Phrasing
  - Minimize the Risk of Bias
  - Maximize the Utility of Response Categories
  - Avoid Making Either Disagreement or Agreement Disagreeable
  - Minimize Fence-Sitting and Floating
- Combining Questions in Indexes
- Designing Questionnaires
  - Build on Existing Instruments
  - Refine and Test Questions
  - Add Interpretive Questions
- Careers and Research
  - Maintain Consistent Focus
- Research in the News: Social Interaction Critical for Mental and Physical Health
  - Order the Questions
  - Make the Questionnaire Attractive
  - Consider Translation
- Organizing Surveys
  - Mailed, Self-Administered Surveys
  - Group-Administered Surveys
  - Telephone Surveys
    - *Reaching Sample Units*
    - *Maximizing Response to Phone Surveys*
  - In-Person Interviews
    - *Balancing Rapport and Control*
    - *Maximizing Response to Interviews*
  - Web Surveys
  - Mixed-Mode Surveys
  - A Comparison of Survey Designs
- Ethical Issues in Survey Research

Conclusions

Research That Matters, Questions That Count

Adolescence is the period when many people first develop romantic relationships, and their success in doing so may influence their life course for many years. Does adolescent mental health have an impact on the quantity and quality of subsequent romantic relationships? Which aspects of mental health are consequential, and for which aspects of romantic relationships? Maggie Thorsen and Jennifer Pearce-Morris analyzed data collected in a longitudinal survey of youth to answer these research questions.

The Child and Young Adult Supplement to the National Longitudinal Survey of Youth began in 1986 to survey every 2 years the biological children of women first surveyed in 1979 when they were between 14 and 21 years old (https://www.nlsinfo.org/content/cohorts/nlsy79-children). Thorsen and Pearce-Morris focused on adolescents who were between the ages of 14 and 16 in 2000–2004 and who were then reinterviewed between the ages of 22 and 24 (in 2008–2012). The survey measures they used included indexes to measure depression, self-esteem, mastery, and impulsivity, as well as number of dating partners in late adolescence and their happiness and level of conflict in those relationships. They found that youth with higher mastery, self-esteem, and impulsivity had more romantic dating partners when they got older, whereas those with high levels of depressive symptoms and low mastery experienced more relationship conflict.

1. What questions would you suggest to measure the quantity and quality of adolescents' romantic relationships? Consider our own experiences, those of others, and what you have learned in your courses. Explain your reasoning.
2. Would you suggest including measures of other aspects of mental health than the four listed above? Which ones and why? Do you think the researchers should have taken other variables into account? (Check the article [p. 230] to see what they decided to take into account.)
3. What would be possible advantages and disadvantages of conducting a survey about these issues using in-person interviews, a self-administered paper questionnaire, a phone survey, or a survey on the web?

In this chapter, you will learn how to write survey questions and how to design survey projects. You will also learn more about the National Longitudinal Survey of Youth. By the end of the chapter, you will know about the major challenges involved in survey research and how they can be reduced by adhering to guidelines for writing questions and using survey designs that match a survey's purpose. As you read the chapter, you can extend you understanding by reading the 2016 *Society and Mental Health* article by Maggie Thorsen and Jennifer Pearce-Morris at the *Investigating the Social World* study site and then test yourself by completing the related interactive exercises for Chapter 8 at edge.sagepub.com/schutt9e.

Thorsen, Maggie L. and Jennifer Pearce-Morris. 2016. "Adolescent Mental Health and Dating in Young Adulthood." *Society and Mental Health* 6(3):223–245.

"Greater self-esteem and mastery will be associated with a higher number of romantic partners across the transition to adulthood in addition to greater relationship happiness and lower relationship conflict during young adulthood." That was the first hypothesis in Maggie Thorsen and Jennifer Pearce-Morris's (2016:227) investigation and their literature review demonstrates its consistency with prior research. However, Thorsen and Pearce-Morris's (2016:226) literature review also reveals that previous tests of this hypothesis have mostly used cross-sectional data or longitudinal studies of very limited duration, or have not focused on the transition from adolescence to young adulthood or on multiple aspects of mental health. Data collected in the Child and Young Adult Supplement to the National

Longitudinal Survey of Youth (NLSY) allowed Thorsen and Pearce-Morris to overcome these limitations. This supplement has followed children of women in the 1979 National Longitudinal Survey of Youth for more than 10 years, so far.

I begin this chapter with a brief review of the reasons for using survey methods, but I will then focus attention on the NLSY supplement and use it to illustrate some key features of survey research. Next, I will discuss guidelines for writing survey questions—a concern in every type of survey research. I will then explain the major steps in questionnaire design and discuss the features of five types of surveys, highlighting the unique problems attending each one and suggesting some possible solutions. I will give particular attention to the ways in which new means of communication such as cell phones and the Internet have been changing survey research since the first supplement survey in 1986. I discuss ethics issues in the final section. By the chapter's end, you should be well on your way to becoming an informed consumer of survey reports and a knowledgeable developer of survey designs. As you read the chapter, I also hope that you will reflect on how mental health influences social relations.

Survey Research in the Social Sciences

Survey research involves the collection of information from a sample of individuals through their responses to questions. Thorsen and Pearce-Morris (2016) turned to survey research data for their study of mental health and romantic relationships because it proved to be an efficient method for systematically collecting data from a broad spectrum of individuals, diverse social settings, and multiple years. As you probably have observed, a great many social scientists—as well as newspaper editors, political pundits, government agencies, and marketing gurus—make the same methodological choice. In fact, surveys have become a multibillion-dollar industry in the United States that shapes what we read in the newspapers, see on TV, and find in government reports (Converse 1984; Tourangeau 2004:776).

Attractions of Survey Research

Survey research owes its popularity to three features: versatility, efficiency, and generalizability. Each of these features is changing as a result of new technologies.

Versatility

First, survey methods are versatile. Although a survey is not the ideal method for testing all hypotheses or learning about every social process, a well-designed survey can enhance our understanding of just about any social issue. The National Longitudinal Survey of Youth covered a range of topics, including social relations, education, and health, and there is hardly any other topic of interest to social scientists that has not been studied at some time with survey methods. Politicians campaigning for election use surveys, as do businesses marketing a product, governments assessing community needs, agencies monitoring program effectiveness, and lawyers seeking to buttress claims of discrimination or select favorable juries.

You can see at a glance the range of topics and research questions that have been the focus of surveys by spending a few minutes on the websites of major survey organizations like Gallup (www.gallup.com), the National Opinion Research Center at the University of Chicago (www.norc.org), and the Pew Research Center (www.pewresearch.org). The Roper Center's archives (ropercenter.cornell.edu) can be accessed from your university's library database (if it has subscribed) and contains over 650,000 questions and answers from surveys conducted since 1935 by over 150 survey firms.

Computer technology has made surveys even more versatile. Computers can be programmed so that different types of respondents are asked different questions. Short videos or pictures can be presented to respondents on a computer screen. An interviewer may give respondents a laptop on which to record their answers to sensitive personal questions, such as about illegal activities, so that not even the interviewer will know what they said (Tourangeau 2004:788–794).

Survey research: Research in which information is obtained from a sample of individuals through their responses to questions about themselves or others.

Efficiency

Surveys also are popular because data can be collected from many people at relatively low cost and, depending on the survey design, relatively quickly. John Mirowsky and Catherine

Ross (2003:207) contracted with the Survey Research Laboratory (SRL) of the University of Illinois for their 25-minute 2003 telephone survey of 2,495 adult Americans. SRL estimated that the survey would incur direct costs of $183,000—that's $73.35 per respondent—and take as long as 1 year to complete. Both this cost and the length of time required were relatively high because SRL made special efforts to track down respondents from the first wave of interviews in 1995. One-shot telephone interviews can cost as little as $30 per subject (Ross 1990). Large mailed surveys cost even less, about $10 to $15 per potential respondent, although the costs can increase greatly when intensive follow-up efforts are made. Surveys of the general population using personal interviews are much more expensive, with costs ranging from about $100 per potential respondent, for studies in a limited geographic area, to $300 or more when lengthy travel or repeat visits are needed to connect with respondents (F. Fowler, personal communication, January 7, 1998; see also Dillman 1982; Groves and Kahn 1979). Surveys through the web have become the quickest way to gather survey data, but there are problems with this method, as I will soon discuss.

Surveys are efficient because many variables can be measured without substantially increasing the time or cost. Mailed questionnaires can include as many as 10 pages of questions before respondents begin to balk. In-person interviews can be much longer. For example, the 2016 General Social Survey (GSS) had three versions in English and Spanish that ranged from 211 to 288 pages (although many sections applied only to subsets of respondents) and measured a total of 949 variables for 2,867 cases interviewed in that year (NORC 2016). The upper limit for phone surveys seems to be about 45 minutes.

Of course, these efficiencies can be attained only in a place with a reliable communications infrastructure (Labaw 1980:xiii–xiv). A reliable postal service, which is required for mail surveys, generally has been available in the United States—although residents of the Bronx, New York, found that grievous delays can occur, while a system audit identified general problems in mail delivery that almost rule out mail surveys (Office of Inspector General 2017). The British postal service, the Royal Mail, has been accused of even worse performance: a "total shambles," with mail abandoned in some cases and purposely misdelivered in other cases (Lyall 2004:A4). Phone surveys have been very effective in countries such as the United States, where 96% of households have phones (Tourangeau 2004:777). Also important to efficiency are the many survey research organizations—about 120 academic and nonprofit organizations in the United States—that provide trained staff and proper equipment (Survey Research Laboratory 2008).

Modern information technology has been a mixed blessing for survey efficiency. The Internet makes it easier to survey some populations, but it leaves out important segments. Caller ID and answering machines make it easy to screen out unwanted calls, but these tools also make it harder to reach people in phone surveys. In addition, as discussed in Chapter 5, a growing number of people use only cell phones. As a result, after a long decline to below 5% in 2001, the percentage of U.S. households without landline telephones climbed to 29% by 2011, and then to 40% by 2013 (Christian et al. 2010; McGeeney and Keeter 2014; U.S. Census Bureau 2013) (see Exhibit 8.1). As a result of these changes, survey researchers must spend more time and money to reach potential respondents (Tourangeau 2004:781–782).

Generalizability

Survey methods lend themselves to probability sampling from large populations. Thus, survey research is appealing when sample generalizability is a central research goal. In fact, survey research is often the only means available for developing a representative picture of the attitudes and characteristics of a large population.

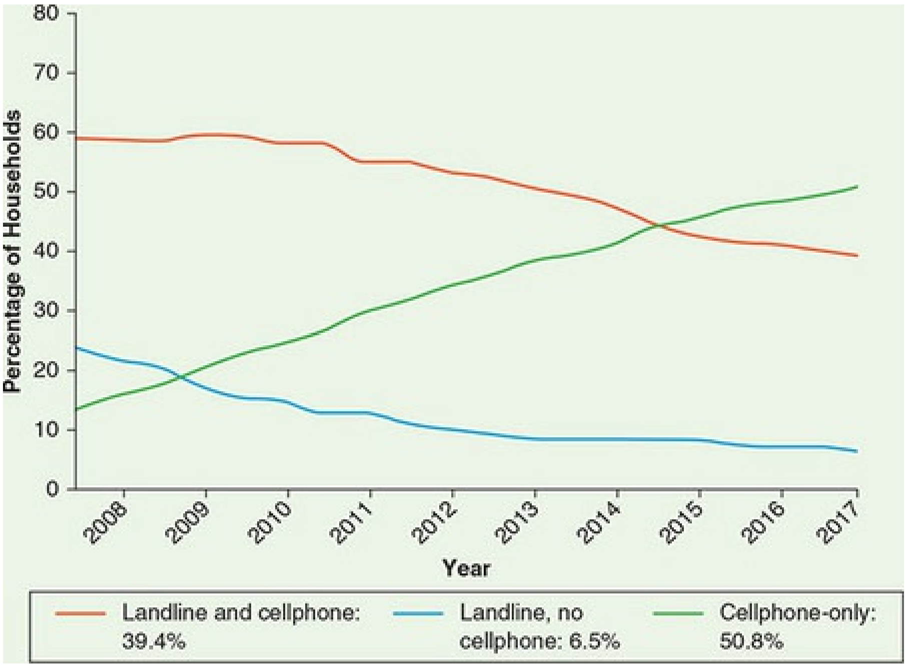
Surveys are also the method of choice when cross-population generalizability is a key concern, because they allow a range of social contexts and subgroups to be sampled. The consistency of relationships can then be examined across the various subgroups. An ambitious Internet-based international survey sponsored by the National Geographic Society (2000) was completed by 80,012 individuals from 178 countries and territories.

Unfortunately (for survey researchers), the new technologies that are lowering the overall rate of response to phone surveys are also making it more difficult to obtain generalizable samples. In 2016, only 13% of households in the United States did not use the Internet at home or work, but in these households persons tend to be elderly, poor, rural, and have no more than a high school education, compared to those who are "connected" (Anderson and Perrin 2016; de Leeuw 2008:321; U.S. Census Bureau 2013).

About 90 percent of U.S. households have a least one cell phone, but among the half of U.S. households that have only cell phone service, the adults are likely to be younger, renters, Hispanic, and poor, while those in households with landline phones only are more likely to elderly (Lavrakas et al. 2017). As a result, although surveys of the general population can include only cell phone numbers, those that target particular subgroups may need to include landlines.

Another challenge in survey research is the growing foreign-born population in the United States (13% in 2014). These individuals often require foreign-language versions of survey forms; otherwise, survey findings may not be generalized to the entire population (Grieco et al. 2012:2; Tourangeau 2004:783; U.S. Census Bureau 2014).

Exhibit 8.1 Percentage of U.S. Households With Cellphones, Landlines, or Both



Source: The Daredevils Without Landlines—And Why Health Experts Are Tracking Them, Alina Selyukh, NPR: All Tech Considered. May 4, 2017. Data source: CDC/NCHS, National Health Interview Survey. Updated May 4, 2017. Alyson Hurt and Alina Selyukh/NPR. Reprinted with permission.

The Omnibus Survey

An omnibus survey shows just how versatile, efficient, and generalizable a survey can be. An omnibus survey covers a range of topics of interest to different social scientists, in contrast to the typical survey that is directed at a specific research question. The omnibus survey has multiple sponsors or is designed to generate data useful to a broad segment of the social science community rather than to answer a particular research question. It is usually directed to a sample of some general population, so the questions, about a range of different issues, are appropriate to at least some sample members.

One of sociology's most successful omnibus surveys is the GSS of the National Opinion Research Center at the University of Chicago. It is an extensive interview administered biennially to a probability sample of at least 2,000 Americans (2,867 in one of three versions in 2016), with a wide range of questions and topic areas chosen by a board of overseers. Some questions are asked of only a randomly selected subset of respondents. This split-ballot design allows more questions without increasing the survey's cost. It also facilitates experiments on the effect of question wording: Different forms of the same question are included in the split-ballot subsets. By 2016, the cumulative GSS database included 5,897 variables. The GSS is widely available to universities, instructors, and students (Davis and Smith 1992; NORC 2011). As the only probability-based in-person interview survey designed to monitor changes in social characteristics and attitudes in the United States, it allows investigation of many social research questions and so has provided the data for 25,000 publications, presentations, and reports (NORC 2016). Many other survey data sets are archived by the Inter-university Consortium for Political and Social Research (ICPSR) (more details about the ICPSR are in Chapter 14).

Omnibus survey: A survey that covers a range of topics of interest to different social scientists.

Split-ballot design: Unique questions or other modifications in a survey administered to randomly selected subsets of the total survey sample, so that more questions can be included in the entire survey or so that responses to different question versions can be compared.

Errors in Survey Research

It might be said that surveys are too easy to conduct. Organizations and individuals often decide that a survey will help solve some important problem because it seems so easy to write up some questions and distribute them. But without careful attention to sampling, measurement, and overall survey design, the effort is likely to be a flop. Such flops are too common for comfort, so the responsible survey researcher must take the time to design surveys properly and to convince sponsoring organizations that this time is worth the effort (Turner and Martin 1984:68).

For a survey to succeed, it must minimize four types of error (Groves 1989:vi, 10–12): (1) poor measurement, (2) nonresponse, (3) inadequate coverage of the population, and (4) sampling error.

Poor measurement.

Measurement error was a key concern in Chapter 4, but there is much more to be learned about how to minimize these errors of observation in the survey process. The theory of satisficing can help us understand the problem. It takes effort to answer survey questions carefully: Respondents have to figure out what each question means, then recall relevant information, and finally decide which answer is most appropriate. Survey respondents satisfice when they reduce the effort required to answer a question by interpreting questions superficially and giving what they think will be an acceptable answer (Krosnick 1999:547–548). A tendency to choose responses appearing earlier in a list of responses—a "primacy effect"—is a similar problem (Toepoel 2016:23). Errors in measurement also arise when respondents are unwilling to disclose their feelings and behaviors, unable to remember past events, and misunderstand survey questions. What people say they can do—such as ability to carry out various tasks—is not necessarily consistent with what they are able to do (Schutt 2011b:88). What people report that they have done is not necessarily what they have actually done (Brenner 2012). A natural desire to say "what the interviewer wants to hear" can generate an "acquiescent response bias" among some respondents, while others may answer questions about sensitive issues in a way they believe is more socially desirable (Toepoel 2016:23).

Presenting clear and interesting questions in a well-organized questionnaire will help reduce measurement error by encouraging respondents to answer questions carefully and to take seriously the request to participate in the survey. Tailoring questions to the specific population surveyed is also important. In particular, persons with less education are more likely to satisfice in response to more challenging questions (Holbrook, Green, and Krosnick 2003; Narayan and Krosnick 1996).

Careful assessment of survey question quality is thus an essential step in survey design. The next section focuses on how to write good survey questions.

Nonresponse.

Nonresponse is a major and growing problem in survey research, although it is a problem that varies between particular survey designs. Social exchange theory can help us understand why nonresponse rates have been growing in the United States and Western Europe since the early 1950s (Dillman 2000:14–15; Groves and Couper 1998:155–189; Tourangeau 2004:782). According to social exchange theory, a well-designed survey effort will maximize the social rewards for survey participation and minimize its costs, as well as establish trust that the rewards will outweigh the costs (Blau 1964). The perceived benefits of survey participation have declined with decreasing levels of civic engagement and with longer work hours (Groves, Singer, and Corning 2000; Krosnick 1999:539–540). Perceived costs have increased with the widespread use of telemarketing and the ability of many people to screen out calls from unknown parties with answering machines and caller ID. In addition, recipients pay for time on cell phone calls, so the ratio of costs to benefits worsens for surveys attempting to reach persons using cell phones (Nagourney 2002). We will review more specifics about nonresponse in this chapter's sections on particular survey methods.

Inadequate coverage of the population.

A poor sampling frame can invalidate the results of an otherwise well-designed survey. We considered the importance of a good sampling frame in Chapter 5; in this chapter, I will discuss special coverage problems related to each of the particular survey methods.

Sampling error.

The process of random sampling can result in differences between the characteristics of the sample members and the population simply on the basis of chance. I introduced this as a topic in Chapter 5. You will learn how to calculate sampling error in Chapter 9.

It is most important to maintain a realistic perspective on the nature of surveys to avoid making unrealistic assumptions about the validity of survey results. Although surveys provide an efficient means for investigating a wide range of issues in large and diverse populations, the data they provide are necessarily influenced by these four sources of error. Survey researchers must make every effort to minimize each one. Only through learning more about different survey features and survey research alternatives can we prepare to weigh the advantages and disadvantages of survey research in particular circumstances and thus assess the value of a survey design in relation to a specific research question.

Writing Survey Questions

Questions are the centerpiece of survey research. Because the way they are worded can have a great effect on the way they are answered, selecting good questions is the single most important concern for survey researchers. All hope for achieving measurement validity is lost unless the questions in a survey are clear and convey the intended meaning to respondents.

You may be thinking that you ask people questions all the time and have no trouble understanding the answers you receive, but can't you also think of times when you've been confused in casual conversation by misleading or misunderstood questions? Now, consider just a few of the differences between everyday conversations and standardized surveys that make writing survey questions much more difficult:

- Survey questions must be asked of many people, not just one.
- The same survey question must be used with each person, not tailored to the specifics of a given conversation.
- Survey questions must be understood in the same way by people who differ in many ways.
- You will not be able to rephrase a survey question if someone doesn't understand it because that would result in a different question for that person.
- Survey respondents don't know you and so can't be expected to share the nuances of expression that help you and your friends and family to communicate.

Writing questions for a particular survey might begin with a brainstorming session or a review of previous surveys. Then, whatever questions are being considered must be systematically evaluated and refined. Although most professionally prepared surveys contain previously used questions as well as some new ones, every question considered for inclusion must be reviewed carefully for its clarity and ability to convey the intended meaning. Questions that were clear and meaningful to one population may not be so to another. Nor can you simply assume that a question used in a previously published study was carefully evaluated.

Adherence to a few basic principles will go a long way toward ensuring clear and meaningful questions. Each of these principles summarizes a great deal of research, although none of them should be viewed as an inflexible mandate (Alwin and Krosnick 1991). As you will learn in the next section, every question must be considered relative to the other questions in a survey. Moreover, every survey has its own unique requirements and constraints; sometimes violating one principle is necessary to achieve others.

Avoid Confusing Phrasing

What's a confusing question? Try this one that I received years ago from the Planetary Society in its National Priorities Survey for the U.S. Space Program:

> The Moon may be a place for an eventual scientific base, and even for engineering resources. Setting up a base or mining experiment will cost tens of billions of dollars in the next century. Should the United States pursue further manned and unmanned scientific research projects on the surface of the Moon?
>
> ☐ Yes ☐ No ☐ No opinion

Does a "yes" response mean that you favor spending tens of billions of dollars for a base or mining experiment? Does "the next century" refer to the 21st century or to the 100 years after the survey (which was distributed in the 1980s)? Could you favor further research projects on the Moon but oppose funding a scientific base or engineering resources? Are engineering resources supposed to have something to do with a mining experiment? Does a mining experiment occur "on the surface of the Moon"? How do you answer if you favor unmanned scientific research projects on the Moon but not manned projects?

There are several ways to avoid such confusing phrasing. In most cases, a simple direct approach to asking a question minimizes confusion. Use shorter rather than longer words and sentences: "brave" rather than "courageous"; "job concerns" rather than "work-related employment issues" (Dillman 2000:52). Try to keep the total number of words to 20 or fewer and the number of commas to 3 or fewer (Peterson 2000:50). However, questions shouldn't be abbreviated in a way that results in confusion: To ask, "In what city or town do you live?" is to focus attention clearly on a specific geographic unit, a specific time, and a specific person (you); the simple format,

> Residential location: ______________________________________,

does not do this.

Sometimes, when sensitive issues or past behaviors are the topic, longer questions can provide cues that make the respondent feel comfortable or aid memory (Peterson 2000:51).

Breaking up complex issues into simple parts also reduces confusion. In a survey about health services, you might be tempted to ask a complex question like this (Schaeffer and Presser 2003):

> During the past 12 months since July 1st, 1987, how many times have you seen or talked with a doctor or a medical assistant about your health? Do not count any times you might have seen a doctor while you were a patient in a hospital, but count all the other times you actually saw or talked to a medical doctor of any kind about your health. (pp. 70–71)

This question can be simplified, thereby reducing confusion, by breaking it up into several shorter questions:

> Have you been a patient in the hospital overnight in the past 12 months since July 1st, 1987?
>
> (Not counting when you were in a hospital overnight) During the past 12 months since July 1st, 1987, how many times did you actually see any medical doctor about your own health?
>
> During the past 12 months since July 1st, 1987, were there any times when you didn't actually see the doctor but saw a nurse or other medical assistant working for the doctor?
>
> During the past 12 months since July 1st, 1987, did you get any medical advice, prescriptions, or results of tests over the telephone from a medical doctor, nurse, or medical assistant working for a doctor? (Cannell et al. 1989:Appendix A, p. 1)

A sure way to muddy the meaning of a question is to use double negatives: "Do you *disagree* that there should *not* be a tax increase?" Respondents have a hard time figuring out which response matches their sentiments. Such errors can easily be avoided with minor wording changes, but even experienced survey researchers can make this mistake unintentionally, perhaps while trying to avoid some other wording problem. For instance, in a survey commissioned by the American Jewish Committee, the Roper polling organization wrote a question about the Holocaust that was carefully worded to be neutral and value free: "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" Among a representative sample of adult Americans, 22% answered that it was possible the extermination never happened (Kifner 1994:A12). Many Jewish leaders and politicians were stunned, wondering how one in five Americans could be so misinformed. But a careful reading of the question reveals how confusing it is: Choosing "possible," the seemingly positive response, means that you don't believe the Holocaust happened. In fact, the Gallup organization then rephrased the question to avoid the double negative, giving a brief definition of the Holocaust and then asking, "Do you doubt that the Holocaust actually happened or not?" Only 9% responded that they doubted it happened. When a wider range of response choices was given, only

2.9% said that the Holocaust “definitely” or “probably” did not happen. To be safe, it’s best just to avoid using negative words such as “don’t” and “not” in questions.

Double negative: A question or statement that contains two negatives, which can muddy the meaning of the question.

So-called double-barreled questions are also guaranteed to produce uninterpretable results because they actually ask two questions but allow only one answer. For example, during the Watergate scandal, Gallup poll results indicated that when the question was “Do you think President Nixon should be impeached and compelled to leave the presidency, or not?” only about a third of Americans supported impeaching President Richard M. Nixon. But when the Gallup organization changed the question to ask respondents if they “think there is enough evidence of possible wrongdoing in the case of President Nixon to bring him to trial before the Senate, or not,” over half answered yes. Apparently, the first, double-barreled version of the question confused support for impeaching Nixon—putting him on trial before the Senate—with concluding that he was guilty before he had had a chance to defend himself (Kagay and Elder 1992:E5).

It is also important to identify clearly what kind of information each question is to obtain. Some questions focus on attitudes, or what people say they want or how they feel. Some questions focus on beliefs, or what people think is true. Some questions focus on behavior, or what people do. And some questions focus on attributes, or what people are like or have experienced (Dillman 1978:79–118; Gordon 1992). Rarely can a single question effectively address more than one of these dimensions at a time.

Whichever type of information a question is designed to obtain, be sure it is asked of only the respondents who may have that information. If you include a question about job satisfaction in a survey of the general population, first ask respondents whether they have a job. You will only annoy respondents if you ask a question that does not apply to them (Schaeffer and Presser 2003:74). These filter questions create skip patterns. For example, respondents who answer *no* to one question are directed to skip ahead to another question, but respondents who answer *yes* go on to the contingent question. Skip patterns should be indicated clearly with an arrow or other mark in the questionnaire, as demonstrated in Exhibit 8.2.

Double-barreled question: A single survey question that actually asks two questions but allows only one answer.

Filter question: A survey question used to identify a subset of respondents who then are asked other questions.

Skip pattern: The unique combination of questions created in a survey by filter questions and contingent questions.

Contingent question: A question that is asked of only a subset of survey respondents.

Minimize the Risk of Bias

Specific words in survey questions should not trigger biases, unless that is the researcher's conscious intent. Biased or loaded words and phrases tend to produce misleading answers. For example, a 1974 survey found that 18% of respondents supported sending U.S. troops "if a situation like Vietnam were to develop in another part of the world." But when the question was reworded to mention sending troops to "stop a communist takeover"—"communist takeover" being a loaded phrase—favorable responses rose to 33% (Schuman and Presser 1981:285).

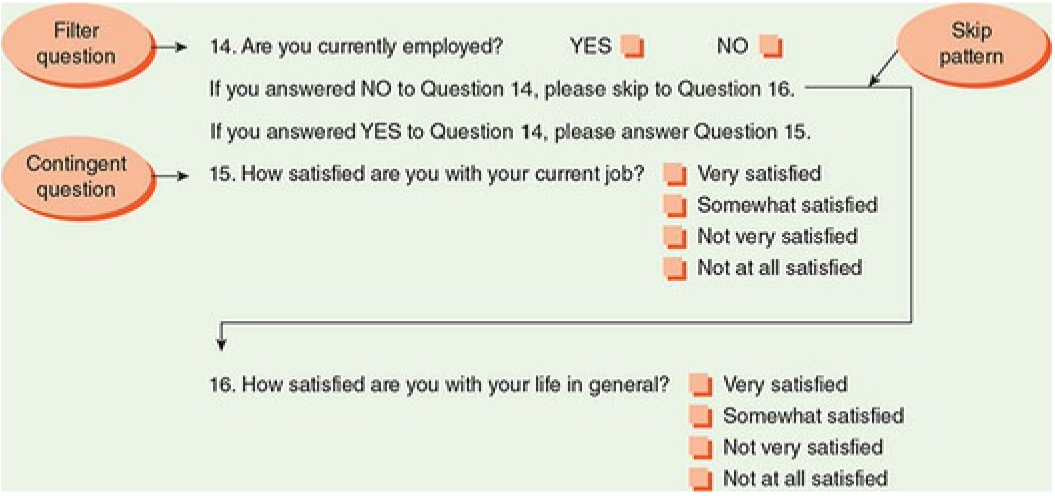
Answers can also be biased by more subtle problems in phrasing that make certain responses more or less attractive to particular groups. To minimize biased responses, researchers have to test reactions to the phrasing of a question. For example, Mirowsky and Ross (personal e-mail, 2009) wanted to ask people, "Do you feel fit?" However, when the University of Illinois Survey Research Laboratory tried out this question with a small sample, people did not seem to understand that they were being asked about their level of energy and general feelings of fitness; they just focused on whether they had some type of health problem. It seemed that people had a biased concept of health as involving only problems rather than including the concept of positive health. As a result, Mirowsky and Ross rephrased the question to be more explicit: "Do you feel physically fit?"

Responses can also be biased when response alternatives do not reflect the full range of possible sentiment on an issue. When people pick a response choice, they seem to be influenced by where they are placing themselves relative to the other response choices. For example, the Detroit Area Study (Turner and Martin 1984:252) asked the following question: "People feel differently about making changes in the way our country is run. In order to keep America great, which of these statements do you think is best?" When the only response choices were "We should be very cautious of making changes" and "We should be free to make changes," only 37% said that we should be free to make changes. However, when a response choice was added that suggested we should "constantly" make changes, 24% picked that response and another 32% chose the "free to make changes" response, for a total of 56% who seemed open to making changes in the way our country is run (Turner and Martin 1984:252). Including the more extreme positive alternative ("constantly" make changes) made the less extreme positive alternative more attractive.

If the response alternatives for a question fall on a continuum from positive to negative, the number of positive and negative categories should be balanced so that one end of the continuum doesn't seem more attractive than the other (Dillman 2000:57–58). If you ask respondents, "How satisfied are you with the intramural sports program here?" and include "completely satisfied" as the most positive possible response, then "completely dissatisfied" should be included as the most negative possible response. This is called a *bipolar* scale. For the same reason, it is also better to state both sides of attitude scales in the question itself: "How satisfied or dissatisfied are you with the intramural sports program here?" (Toepoel 2016:27).

Of course, the advice to minimize the risk of bias is intentionally ignored by those who conduct surveys to elicit bias. This is the goal of *push polling,* a technique that has been used in some political campaigns. In a push poll, the pollsters for a candidate call potential voters and ask them a series of questions that convey negative information about the opposing candidate. It's really not a survey at all—just a propaganda effort—but it casts reputable survey research (and ethical political polling firms) in a bad light (Connolly and Manning 2001).

Exhibit 8.2 Filter Questions and Skip Patterns



14. Are you currently employed? YES ☐ NO ☐

If you answered NO to Question 14, please skip to Question 16.

If you answered YES to Question 14, please answer Question 15.

15. How satisfied are you with your current job?
☐ Very satisfied
☐ Somewhat satisfied
☐ Not very satisfied
☐ Not at all satisfied

16. How satisfied are you with your life in general?
☐ Very satisfied
☐ Somewhat satisfied
☐ Not very satisfied
☐ Not at all satisfied

Maximize the Utility of Response Categories

Response choices should be considered carefully because they help respondents to understand what the question is about and what types of responses are viewed as relevant (Clark and Schober 1994). Questions with fixed response choices must provide one and only one possible response for everyone who is asked the question—that is, the response choices must be exhaustive and mutually exclusive. Ranges of ages, incomes, years of schooling, and so forth should not overlap and should provide a response option for all respondents.

There are two exceptions to this principle: (1) Filter questions may tell some respondents to skip over a question (the response choices do not have to be exhaustive), and (2) respondents may be asked to “check all that apply” (the response choices are not mutually exclusive). Even these exceptions should be kept to a minimum. Respondents to a self-administered paper questionnaire should not have to do a lot of skipping around, or they may lose interest in answering carefully all the applicable questions. Some survey respondents react to a “check all that apply” request by just checking enough responses so that they feel they have “done enough” for that question and then ignoring the rest of the choices (Dillman 2000:63).

Vagueness in the response choices is also to be avoided. Questions about thoughts and feelings will be more reliable if they refer to specific times or events (Turner and Martin 1984:300). Usually a question like “On how many days did you read the newspaper in the last week?” produces more reliable answers than one like “How often do you read the newspaper?” in which response choices of “frequently,” “sometimes,” and “never” are given (Toepoel 2016:27). In their survey, Mirowsky and Ross (2001:2) sensibly asked the question “Do you currently smoke 7 or more cigarettes a week?” rather than the vaguer question “Do you smoke?” Of course, being specific doesn’t help if you end up making unreasonable demands of your respondents’ memories. One survey asked, “During the past 12 months, about how many times did you see or talk to a medical doctor?” According to their written health records, respondents forgot 60% of their doctor visits (Goleman 1993b:C11). So unless your focus is on major or routine events that are unlikely to have been forgotten, limit questions about specific past experiences to the past month.

Another problem to avoid is making fine distinctions at one end of a set of response choices, while using broader categories at the other end. For example, in response to the question “How many hours per day do you typically watch TV?” 78% said they watched 2 1 2 hours or less when the first five response categories distinguished five response choices from “ 1 2 hour or less” to “ 2 –2 1 2 hours,” but when the first response choice was “ 2 1 2 hours or less,” only 46.4% picked it (Toepoel 2016:20–21).

Sometimes, problems with response choices can be corrected by adding questions. For example, if you ask, “How many years of schooling have you completed?” someone who dropped out of high school but completed the requirements for a General Equivalency Diploma (GED) might not be sure how to respond. By asking a second question, “What is the highest degree you have received?” you can provide the correct alternative for those with a GED as well as for those who graduated from high school.

Adding questions may also improve memory about specific past events. Imagine the problem you might have answering the question “How often did you receive help from classmates while preparing for exams or completing assignments during the last month? (very often, somewhat often, occasionally, rarely, or never).” Now, imagine a series of questions that asks you to identify the exams and assignments you had in the past month and, for each one, inquires whether you received each of several types of help from classmates: study suggestions, study sessions, related examples, general encouragement, and so on. The more specific focus on particular exams and assignments should result in more complete recall (Dykema and Schaeffer 2000).

Response choices should be matched to the question they follow and reflect meaningful distinctions, as well as cover the range of possible responses—another way of saying that the response choices should be mutually exclusive and exhaustive. If the question is “How satisfied are you with your job?” the response choices should focus on distinctions between levels of satisfaction and might range from “very satisfied” to “somewhat,” “not very,” and “not at all satisfied.” If one response choice is “somewhat satisfied,” “moderately satisfied” would not be a good additional response because it does not reflect a meaningful distinction. When measuring the importance of something to respondents, response choices should include “extremely important” as well as “not at all important,” “slightly important,” “moderately important,” and “very important” because people tend to rank many issues as “very important.”

One common approach for measures of attitude intensity is to present a statement and then ask respondents to indicate their degree of agreement or disagreement. The last question in this section, about “my misfortunes,” is an example, using the form known as a Likert item (after social psychologist Rensis Likert, who popularized this approach). A Likert item phrases an attitude in terms of one end of a continuum, so that the responses ranging from “strongly agree” to “strongly disagree” cover the full range of possible agreement. However, the risk of agreement bias should be considered carefully when interpreting responses to Likert-style items (see the next section).

Other words used to distinguish points on an ordinal scale of attitude intensity are reflected in the response choices in Exhibit 8.3. One important decision is whether to use unipolar distinctions, such as “not at all” to “extremely,” or bipolar distinctions, such as “very comfortable” to “very uncomfortable.” The advantages of using bipolar response options are discussed in the next section. How many response categories are desirable? Five categories work well for unipolar ratings, and seven will capture most variation on bipolar ratings (Krosnick 2006; Schaeffer and Presser 2003:78–79). Responses are more reliable when these categories are labeled (labeled unipolar response options) rather than identified only by numbers (unlabeled unipolar response options) (Krosnick 1999:544; Schaeffer and Presser 2003:78). Exhibit 8.3 shows these alternatives, based on a question and response alternatives used in a survey about education and health (Mirowsky and Ross 2001). A special consideration for ratings is whether to include a middle, neutral response option. This issue is discussed later.

Avoid Making Either Disagreement or Agreement Disagreeable

People often tend to "agree" with a statement just to avoid seeming disagreeable. This is termed *agreement bias,* social desirability bias, or an *acquiescence effect.* You can see the impact of this human tendency in a 1974 University of Michigan Survey Research Center survey that asked who was to blame for crime and lawlessness in the United States (Schuman and Presser 1981:208). When one question stated that individuals were more to blame than social conditions, 60% of the respondents agreed. But when the question was rephrased and respondents were asked, in a balanced fashion, whether individuals or social conditions were more to blame, only 46% chose individuals. Numerous studies of agreement bias suggest that about 10% of respondents will "agree" just to be agreeable, without regard to what they really think (Krosnick 1999:553).

Likert item: A statement followed by response choices ranging from "strongly agree" to "strongly disagree."

Bipolar response options: Response choices to a survey question that include a middle category and parallel responses with positive and negative valence (can be labeled or unlabeled).

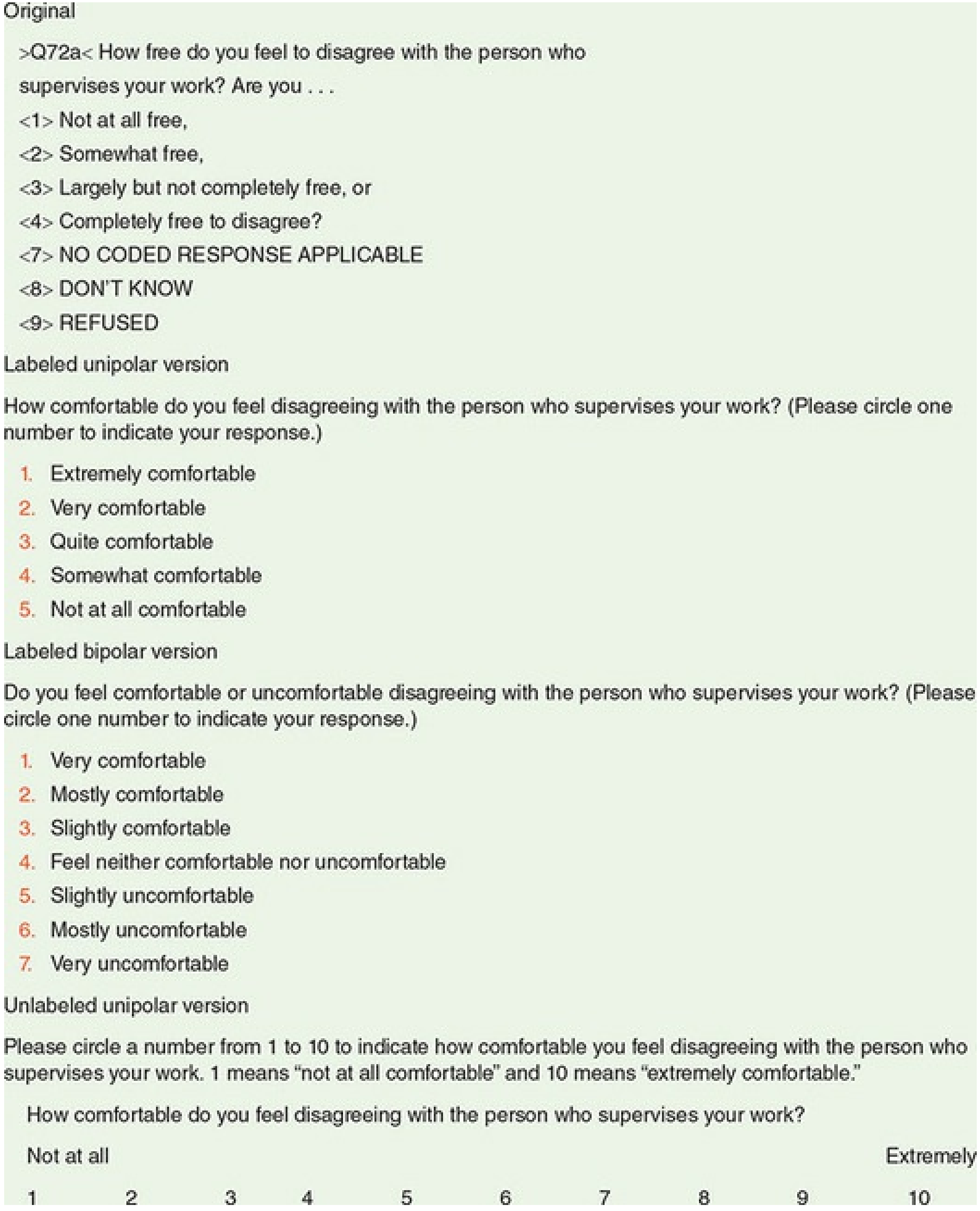
Labeled unipolar response options: Response choices for a survey question that use words to identify categories ranging from low to high (or high to low).

Unlabeled unipolar response options: Response choices for a survey question that use numbers to identify categories ranging from low to high (or high to low).

Social desirability bias: The tendency to "agree" with a statement just to avoid seeming disagreeable.

You can take several steps to reduce the likelihood of agreement bias. As a general rule, you should present both sides of attitude scales in the question itself (Dillman 2000:61–62): "In general, do you believe that *individuals* or *social conditions* are more to blame for crime and lawlessness in the United States?" The response choices themselves should be phrased to make each one seem as socially approved, as "agreeable," as the others. You should also consider replacing a range of response alternatives that focus on the word *agree* with others. For example, "To what extent do you support or oppose the new health care plan?" (response choices range from "strongly support" to "strongly oppose") is probably a better approach than the question "To what extent do you agree or disagree with the statement: 'The new health care plan is worthy of support'?" (response choices range from "strongly agree" to "strongly disagree"). For the same reason, simple true–false and yes–no response choices should be avoided (Schaeffer and Presser 2003:80–81).

Exhibit 8.3 Labeled Unipolar, Unlabeled Unipolar, and Bipolar Response Options



Original

>Q72a< How free do you feel to disagree with the person who supervises your work? Are you . . .

<1> Not at all free,
<2> Somewhat free,
<3> Largely but not completely free, or
<4> Completely free to disagree?
<7> NO CODED RESPONSE APPLICABLE
<8> DON'T KNOW
<9> REFUSED

Labeled unipolar version

How comfortable do you feel disagreeing with the person who supervises your work? (Please circle one number to indicate your response.)

1. Extremely comfortable
2. Very comfortable
3. Quite comfortable
4. Somewhat comfortable
5. Not at all comfortable

Labeled bipolar version

Do you feel comfortable or uncomfortable disagreeing with the person who supervises your work? (Please circle one number to indicate your response.)

1. Very comfortable
2. Mostly comfortable
3. Slightly comfortable
4. Feel neither comfortable nor uncomfortable
5. Slightly uncomfortable
6. Mostly uncomfortable
7. Very uncomfortable

Unlabeled unipolar version

Please circle a number from 1 to 10 to indicate how comfortable you feel disagreeing with the person who supervises your work. 1 means "not at all comfortable" and 10 means "extremely comfortable."

How comfortable do you feel disagreeing with the person who supervises your work?

Not at all Extremely
1 2 3 4 5 6 7 8 9 10

Source: Based on Mirowsky and Ross (2001:9).

You may also gain a more realistic assessment of respondents' sentiment by adding to a question a counterargument in favor of one side to balance an argument in favor of the other side. Thus, don't just ask in an employee survey whether employees should be required to join the union; instead, ask whether employees should be required to join the union or be able to make their own decision about joining. In one survey, 10% more respondents said they favored mandatory union membership when the counterargument was left out than when it was included. It is reassuring to know, however, that this approach does not change the distribution of answers to questions about which people have very strong beliefs (Schuman and Presser 1981:186).

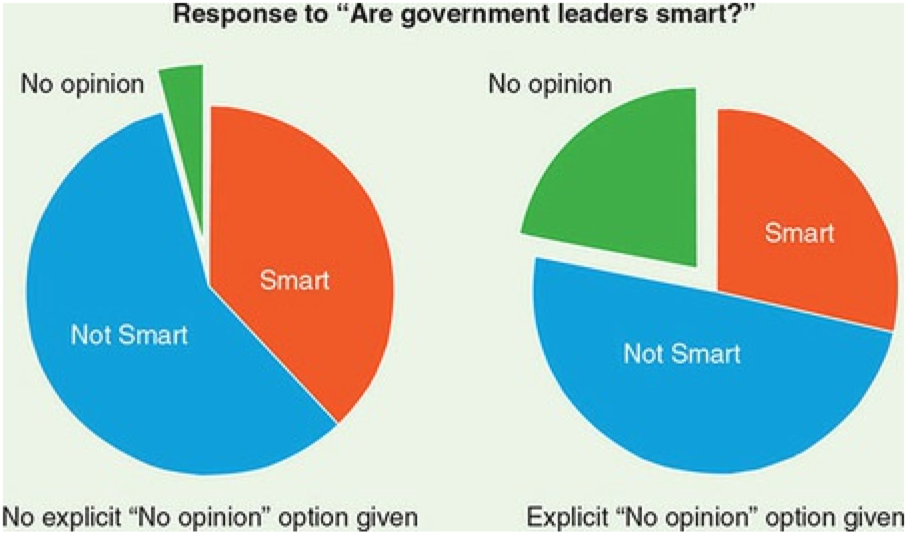
When an illegal or socially disapproved behavior or attitude is the focus, we have to be concerned that some respondents will be reluctant to agree that they have ever done or thought such a thing. In this situation, the goal is to write a question and response choices that make agreement seem more acceptable. For example, Dillman (2000:75) suggests that we ask, “Have you ever taken anything from a store without paying for it?” rather than “Have you ever shoplifted something from a store?” Asking about a variety of behaviors or attitudes that range from socially acceptable to socially unacceptable will also soften the impact of agreeing with those that are socially unacceptable.

Minimize Fence-Sitting and Floating

Two related problems in writing survey questions also stem from people's desire to choose an acceptable answer. There is no uniformly correct solution to these problems; researchers have to weigh the alternatives in light of the concept to be measured and whatever they know about the respondents.

Fence-sitters, people who see themselves as being neutral, may skew the results if you force them to choose between opposites. In most cases, about 10% to 20% of such respondents —those who do not have strong feelings on an issue—will choose an explicit middle, neutral alternative (Schuman and Presser 1981:161–178). Having an explicit neutral response option is generally a good idea: It identifies fence-sitters and tends to increase measurement reliability (Schaeffer and Presser 2003:78).

Exhibit 8.4 The Effect of Floaters on Public Opinion Polls



Source: Based on Schuman and Presser (1981:121).

Even more people can be termed floaters: respondents who choose a substantive answer when they really don't know or have no opinion. A third of the public will provide an opinion on a proposed law that they know nothing about if they are asked for their opinion in a closed-ended survey question that does not include "Don't know" as an explicit response choice. However, 90% of these persons will select the "Don't know" response if they are explicitly given that option. On average, offering an explicit response option increases the "Don't know" responses by about a fifth (Schuman and Presser 1981:113–160).

Fence-sitters: Survey respondents who see themselves as being neutral on an issue and choose a middle (neutral) response that is offered.

Floaters: Survey respondents who provide an opinion on a topic in response to a closed-ended question that does not include a "Don't know" option, but who will choose "Don't know" if it is available.

Exhibit 8.4 depicts the results of one study that tested the effect of giving respondents an explicit "No opinion" option to the question "Are government leaders smart?" Notice how many more people chose "No opinion" when they were given that choice than when their only explicit options were "Smart" and "Not smart."

Despite the prevalence of floating, people often have an opinion but are reluctant to express it. Actually, most political pollsters use forced-choice questions without a "Don't know" option. Just after President Bill Clinton's victory, Frank Newport, editor in chief of the Gallup poll, defended pollsters' efforts to get all prospective voters to declare a preferred candidate:

> It would not be very instructive for pollsters . . . to allow large numbers of voters to claim they are undecided all through the election season. We would miss the dynamics of change, we would be unable to tell how well candidates were doing in response to events, and publicly released polls would be out of synchronization with private, campaign polls. (Newport 1992:A28)

Because there are so many floaters in the typical survey sample, the decision to include an explicit "Don't know" option for a question is important. Unfortunately, the inclusion of an explicit "Don't know" response choice leads some people who do have a preference to take the easy way out—to satisfice—and choose "Don't know." This is particularly true in surveys of less-educated populations—except for questions that are really impossible to decipher, to which more educated persons are likely to say they "don't know" (Schuman and Presser 1981:113–146). As a result, survey experts now recommend that questions not include "Don't know" or "No opinion" options (Krosnick 1999:558; Schaeffer and Presser 2003:80). Adding an open-ended question in which respondents are asked to discuss their opinions can help identify respondents who are floaters (Smith 1984).

Researchers who use in-person or telephone interviews (rather than self-administered questionnaires) may get around the dilemma somewhat by reading the response choices without a middle or "Don't know" alternative but recording a noncommittal response if it is offered. Mirowsky and Ross's (2001) questionnaire for their phone survey about education and health included the following example (responses in ALL CAPS were not read):

My misfortunes are the result of mistakes I have made.

(Do you . . .)

1. <1> Strongly agree,
2. <2> Agree,
3. <3> Disagree, or
4. <4> Strongly disagree?
5. <7> NO CODED RESPONSE APPLICABLE
6. <8> DON'T KNOW
7. <9> REFUSED

Combining Questions in Indexes

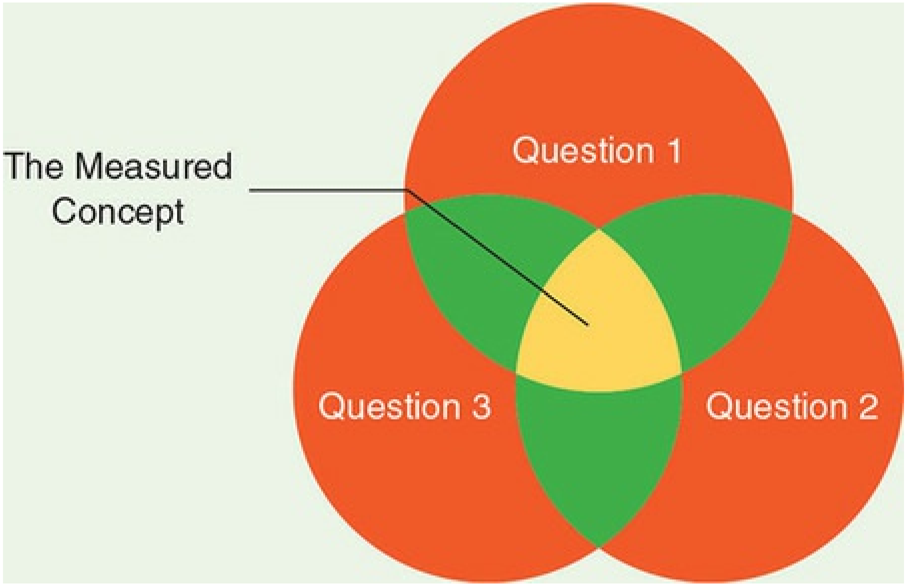
Writing single questions that yield usable answers is always a challenge. Simple though they may seem, single questions are prone to error because of idiosyncratic variation, which occurs when individuals' responses vary because of their reactions to particular words or ideas in the question. Differences in respondents' backgrounds, knowledge, and beliefs almost guarantee that some will understand the same question differently.

Forced-choice questions: Closed-ended survey questions that do not include "Don't know" as an explicit response choice.

Idiosyncratic variation: Variation in responses to questions that is caused by individuals' reactions to particular words or ideas in the question instead of by variation in the concept that the question is intended to measure.

In some cases, the effect of idiosyncratic variation can be dramatic. For example, when people were asked in a survey whether they would "forbid" public speeches against democracy, 54% agreed. When the question was whether they would "not allow" public speeches against democracy, 75% agreed (Turner and Martin 1984:chap. 5). Respondents are less likely to respond affirmatively to the question "Did you see a broken headlight?" than they are to the question "Did you see *the* broken headlight?" (Turner and Martin 1984:chap. 9).

Exhibit 8.5 Overlapping Dimensions of a Concept



The guidelines in this chapter for writing clear questions should help reduce idiosyncratic variation caused by different interpretations of questions. But the best option is often to develop multiple questions about a concept and then to average the responses to those questions in a composite measure termed an *index* or *scale.*

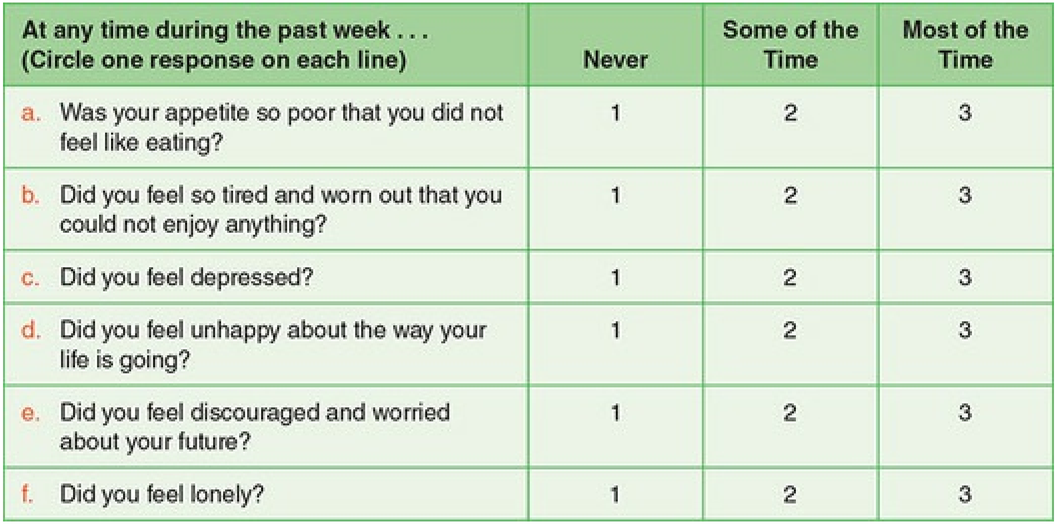
The idea is that idiosyncratic variation in response to particular questions will average out, so that the main influence on the combined measure will be the concept upon which all the questions focus. The index can be considered a more complete measure of the concept than can any one of the component questions.

Creating an index is not just a matter of writing a few questions that seem to focus on a concept. Questions that seem to you to measure a common concept might seem to respondents to concern several different issues. The only way to know that a given set of questions does, in fact, form an index is to administer the questions to people like those you plan to study. If a common concept is being measured, people's responses to the different questions should display some consistency. In other words, responses to the different questions should be correlated. Exhibit 8.5 illustrates an index in which responses to the items are correlated; the substantial area of overlap indicates that the questions are measuring a common concept. Special statistics called *reliability measures* help researchers decide whether responses are consistent. The most common of these, Cronbach's alpha (α), varies from 0 to 1. A score of 0 indicates that answers to different questions in the index are completely unrelated, while a score of 1 indicates that the same response is given to every question in the index. An index is not considered sufficiently reliable unless α is at least .7. When an index score is used to make critical decisions, such as about employment or academic placement, an even higher level of reliability should be required (DeVellis 2017:146).

Because of the popularity of survey research, indexes already have been developed to measure many concepts, and some of these indexes have proved to be reliable in a range of studies. It usually is much better to use such an index to measure a concept than to try to devise questions to form a new index. Use of a preexisting index both simplifies the work involved in designing a study and facilitates comparison of findings to those obtained in other studies.

The questions in Exhibit 8.6 are a different form of the index to measure the concept of depression—the Center for Epidemiologic Studies Depression Index (CES-D)—that was used in the National Longitudinal Survey of Youth (see "Research That Matters" at the beginning of this chapter). Many researchers in different studies have found that these questions form a reliable index. Note that each question concerns a symptom of depression. People may have idiosyncratic reasons for having a particular symptom without being depressed; for example, persons who have been suffering a physical ailment may report that they have a poor appetite. But by combining the answers to questions about several symptoms, the index score reduces the impact of this idiosyncratic variation.

Exhibit 8.6 Example of an Index: Short Form of the Center for Epidemiologic Studies Depression Index (CES-D)



At any time during the past week . . . (Circle one response on each line)	Never	Some of the Time	Most of the Time
a. Was your appetite so poor that you did not feel like eating?	1	2	3
b. Did you feel so tired and worn out that you could not enjoy anything?	1	2	3
c. Did you feel depressed?	1	2	3
d. Did you feel unhappy about the way your life is going?	1	2	3
e. Did you feel discouraged and worried about your future?	1	2	3
f. Did you feel lonely?	1	2	3

Source: Radloff, Lenore. 1977. “The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.” Applied Psychological Measurement 1:385–401.

Three cautions are in order:

1. *Our presupposition that each component question is indeed measuring the same concept may be mistaken.* Although we may include multiple questions in a survey to measure one concept, we may find that answers to the questions are not related to one another, and so the index cannot be created. Alternatively, we may find that answers to just a few of the questions are not related to the answers given to most of the others. We may therefore decide to discard these particular questions before computing the average that makes up the index.
2. *Combining responses to specific questions can obscure important differences in meaning among the questions.* My research on the impact of AIDS prevention education in shelters for the homeless provides an example. In this research, I asked a series of questions to ascertain respondents’ knowledge about HIV risk factors and about methods of preventing exposure to those risk factors. I then combined these responses into an overall knowledge index. I was somewhat surprised to find that the knowledge index scores were no higher in a shelter with an AIDS education program than in a shelter without such a program. However, further analysis showed that respondents in the shelter with an AIDS education program were more knowledgeable than the other respondents about the specific ways of preventing AIDS, which were, in fact, the primary focus of the program. Combining responses to these questions with the others about general knowledge of HIV risk factors obscured an important finding (Schutt, Gunston, and O’Brien 1992).
3. *The questions in an index may cluster together in subsets.* All the questions may be

> measuring the intended concept, but we may conclude that this concept actually has several different aspects. A multidimensional index has been obtained. This conclusion can, in turn, help us refine our understanding of the original concept. For example, Carlo DiClemente and colleagues (1994) sought to determine how confident individuals in treatment for alcoholism were that they could abstain from drinking in different situations that presented typical drinking cues. The 20 situations they presented were of four different types: (1) negative affect, (2) social/positive, (3) physical and other concerns, or (4) withdrawal and urges. The questions used to measure these different dimensions are mixed together in the Alcohol Abstinence Self-Efficacy Scale so that individuals completing the index may not be aware of them (see Exhibit 8.7). However, the answers to questions representing the particular dimensions tend to be more similar to each other than to answers to questions representing other dimensions—they tend to cluster together. By creating subscales for each of these dimensions, researchers can identify not only the level of confidence in ability to resist drinking cues (abstinence self-efficacy) but also the types of drinking cues that are most difficult for individuals to resist.

Differences in wording can also result in responses to subsets of questions in an index clustering together. To encourage respondents to read each question in an index, rather than checking off their answers without thinking about the wording based on a consistent response set, some researchers reverse the wording of some questions in an index so that earlier responses refer to lower levels of the concept (while other questions are phrased so that earlier responses refer to higher levels of the concept). For example, in the next statement, from the NLSY (2014 Child Self-Administered supplement), the response of "Strongly Agree" supports treating girls and boys the same, while the same response to the following statement indicates support of unequal treatment:

How much do you agree or disagree with the following statements?

Girls and boys should be treated the same at school.

1. Strongly Agree
2. Agree
3. Disagree
4. Strongly Disagree

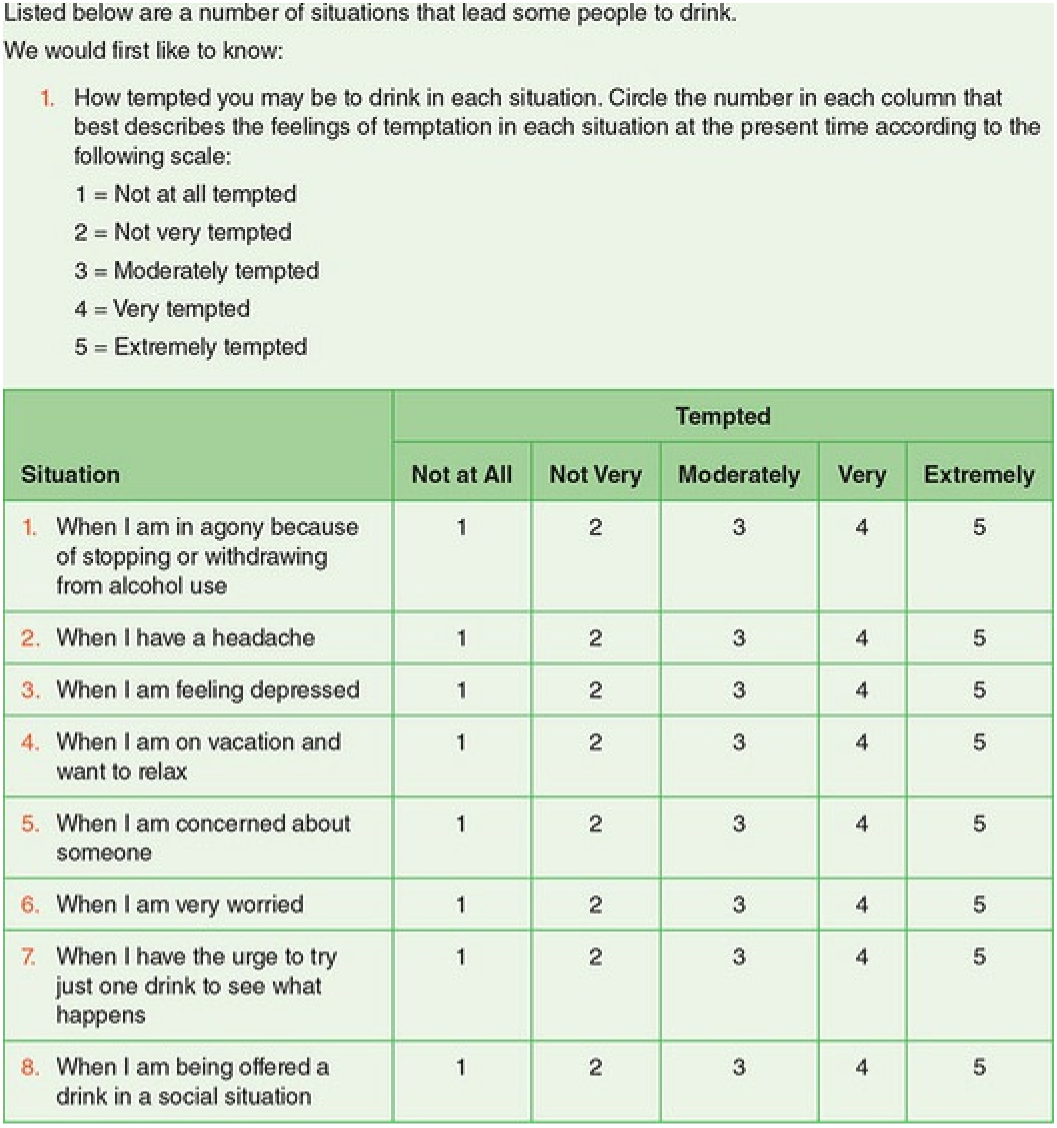
A girl should NOT let a boy know she is smarter than he is.

1. Strongly Agree
2. Agree
3. Disagree
4. Strongly Disagree

The numbers representing the responses to one subset are then reverse scored before the index is created, so that a higher numerical value always means a higher value on the concept. Unfortunately, the responses to the positively and negatively worded questions will tend to cluster together even though the statements themselves are phrased in opposing ways. As a result, the reliability of the overall index will be lower than if all the questions had been worded in the same direction in the first place.

An index score is usually calculated as the arithmetic average or sum of responses to the component questions, so that every question that goes into the index counts equally. Exhibit 8.7 shows how an index score is calculated from answers to the questions in the Alcohol Abstinence Self-Efficacy Scale (AASE). The interitem reliability of an index (Cronbach's ɑ) will increase with the number of items included in the index, even when the association between the individual items stays the same.

Exhibit 8.7 Alcohol Abstinence Self-Efficacy Scale (AASE)



Listed below are a number of situations that lead some people to drink.

We would first like to know:

1. How tempted you may be to drink in each situation. Circle the number in each column that best describes the feelings of temptation in each situation at the present time according to the following scale:

   1 = Not at all tempted
   2 = Not very tempted
   3 = Moderately tempted
   4 = Very tempted
   5 = Extremely tempted

	Tempted				
Situation	**Not at All**	**Not Very**	**Moderately**	**Very**	**Extremely**
1. When I am in agony because of stopping or withdrawing from alcohol use	1	2	3	4	5
2. When I have a headache	1	2	3	4	5
3. When I am feeling depressed	1	2	3	4	5
4. When I am on vacation and want to relax	1	2	3	4	5
5. When I am concerned about someone	1	2	3	4	5
6. When I am very worried	1	2	3	4	5
7. When I have the urge to try just one drink to see what happens	1	2	3	4	5
8. When I am being offered a drink in a social situation	1	2	3	4	5

Source: Journal of Studies on Alcohol, volume 55, pp 141–148, 1994. Center for Alcohol Studies, Rutgers.

Another approach to creating an index score is to give different weights to the responses to different questions before summing or averaging the responses. Such a weighted index is also termed a *scale.* The scaling procedure might be as simple as arbitrarily counting responses to one question as worth two or three times as much as responses to another question, but most often, the weight applied to each question is determined through empirical testing. For example, based on Christopher Mooney and Mei Hsien Lee's (1995) research on abortion law reform, the scoring procedure for a scale of support for abortion might give a 1 to agreement that abortion should be allowed "when the pregnancy resulted from rape or incest" and a 4 to agreement with the statement that abortion should be allowed "whenever a woman decided she wanted one." In other words, agreeing that abortion is allowable in any circumstances is much stronger support for abortion rights than is agreeing that abortion should be allowed in the case of rape or incest.

Designing Questionnaires

Survey questions are answered as part of a questionnaire (or interview schedule, as it's often called in interview-based studies), not in isolation from other questions. The context created by the questionnaire has a major impact on how individual questions are interpreted and whether they are even answered. As a result, survey researchers must give very careful attention to the design of the questionnaire as well as to the individual questions that it includes.

The way a questionnaire should be designed varies with the specific survey method used and with other particulars of a survey project. There can be no precise formula for identifying questionnaire features that reduce error. Nonetheless, some key principles should guide the design of any questionnaire, and some systematic procedures should be considered for refining it. I will use Mirowsky and Ross's (1999) questionnaire for studying the psychological effects of changes in a household structure to illustrate some of these principles and procedures.

Build on Existing Instruments

If another researcher already has designed a set of questions to measure a key concept, and evidence from previous surveys indicates that this measure is reliable and valid, then, by all means, use that instrument. Resources such as Delbert Miller and Neil J. Salkind's (2002) *Handbook of Research Design and Social Measurement* can give you many ideas about existing instruments; your literature review at the start of a research project should be an even better source.

But there is a trade-off here. Questions used previously may not concern quite the right concept or may not be appropriate in some ways to your population. For example, *sense of control* was a key concept in a Mirowsky and Ross survey on aging and sense of control (Mirowsky 1999:13), so they carefully reviewed prior research that had measured this concept. They found that people who were older and had lower incomes tended to "agree" more with statements to which they were asked to respond. As a result, Mirowsky and Ross decided to use a measure of sense of control that was not subject to agreement bias. A good rule of thumb is to use a previously designed instrument if it measures the concept of concern to you and if you have no clear reason for thinking that the instrument is not appropriate with your survey population. Before making a final decision, you should ask other researchers for their opinions about concepts that are difficult to measure.

Refine and Test Questions

Adhering to the preceding question-writing guidelines will go a long way toward producing a useful questionnaire. However, simply asking what appear to you to be clear questions does not ensure that people have a consistent understanding of what you are asking. You need some external feedback—the more of it, the better. This feedback is obtained from some type of pretest (Dillman 2000:140–147). Pretesting is an essential step in preparing any survey.

Questionnaire: The survey instrument containing the questions in a self-administered survey.

Interview schedule: The survey instrument containing the questions asked by the interviewer in an in-person or phone survey.

One important form of feedback results from simply discussing the questionnaire content with others. Persons who should be consulted include expert researchers, key figures in the locale or organization to be surveyed (e.g., elected representatives, company presidents, and community leaders), and individuals from the population to be sampled. Run your list of variables and specific questions by such figures whenever you have a chance. Reviewing the relevant literature to find results obtained with similar surveys and comparable questions is also an important step to take, if you haven't already conducted such a review before writing your questions. Forming a panel of experts to review the questions can also help: Stanley Presser and Johnny Blair (1994) recommend a panel of a psychologist, a questionnaire design expert, and a general methodologist (cited in Peterson 2000:116).

Another increasingly popular form of feedback comes from guided discussions between potential respondents, called *focus groups,* to check for consistent understanding of terms and to identify the range of events or experiences about which people will be asked to report. By listening to and observing the focus group discussions, researchers can validate their assumptions about what level of vocabulary is appropriate and what people are going to be reporting (Fowler 1995). (See Chapter 10 for more about this technique.)

Professional survey researchers also use a technique for improving questions called the cognitive interview (Dillman 2000:66–67; Fowler 1995). Although the specifics vary, the basic approach is to ask people to describe what they are thinking when they answer questions. The researcher asks a test question, then probes with follow-up questions about how the respondent understood one or more words in the question, how confusing it was, and so forth (Schaeffer and Presser 2003:82). This method can identify many problems with proposed questions, particularly if the individuals interviewed reflect the population to be surveyed. Different particular approaches to cognitive interviewing can identify different problems with survey questions. However, there is as yet no single approach to cognitive interviewing that can be considered most effective (Presser et al. 2004:109–130).

In a traditional survey pretest, interviewers administer the questionnaire to a small set of respondents (perhaps 15–25) who are similar to those who will be sampled in the planned survey. After the interviews are completed, the interviewers discuss the experience with the researcher and, through this discussion, try to identify questions that caused problems. Try it yourself if you develop a questionnaire. Prepare for the pretest by completing the questionnaire yourself and then revising it. Next, try it out on some colleagues or other friends, and revise it again. For the actual pretest, draw a small sample of individuals from the population you are studying, or one very similar to it, and try out the survey procedures with them, including as many mailings as you plan if you will mail your questionnaire, and actual interviews if you plan to conduct in-person interviews.

In the pretest version of a written questionnaire, you may include some space for individuals to comment on each key question or, with in-person interviews, audio-record the test interviews for later review. You can also check questions in the pretest for their susceptibility to social desirability bias by including a short index of social desirability (Strahan and Berbasi 1972). Consider excluding survey questions that have a strong social desirability component (DeVellis 2017:136).

Conclude the pretest by reviewing the responses to each question and listening to the audio-recordings and reading the comments. Revise any questions that respondents do not seem to interpret as you had intended or that are not working well for other reasons. If the response rate is relatively low, consider whether it can be improved by some modifications in procedures.

The value of a pretest can be enhanced with behavior coding, in which a researcher observes the interviews or listens to recorded interviews and codes, according to strict rules, the number of times that difficulties occur with questions (Krosnick 1999:541). Such difficulties include respondents asking for clarification and interviewers rephrasing questions rather than reading them verbatim (Presser and Blair 1994:74–75). This information is then used to improve question wording and instructions for interviewers about reading the questions (Schaeffer and Presser 2003:82).

Which method of improving questions is best? Each has unique advantages and disadvantages. Behavior coding, with its clearly specified rules, is the most reliable method across interviewers and repetitions, whereas simple pretesting is the least reliable. However, focus groups or cognitive interviews are better for understanding the bases of problems with particular questions. Review of questions by an expert panel is the least expensive method and identifies the greatest number of problems with questions (Presser and Blair 1994).

Cognitive interview: A technique for evaluating questions in which researchers ask people test questions and then probe with follow-up questions to learn how they understood the question and what their answers mean.

Survey pretest: A method of evaluating survey questions and procedures by testing them on a small sample of individuals like those to be included in the actual survey and then reviewing responses to the questions and reactions to the survey procedures.

Behavior coding: An observation in which the researcher categorizes, according to strict rules, the number of times certain behaviors occur.

Add Interpretive Questions

A survey researcher can also include interpretive questions in the survey itself to help the researcher understand what the respondent meant by his or her responses to particular questions. An example from a study of people with motor vehicle driving violations illustrates the importance of interpretive questions:

> When asked whether their emotional state affected their driving at all, respondents would reply that their emotions had very little effect on their habits. Then, when asked to describe the circumstances surrounding their last traffic violation, respondents typically replied, "I was mad at my girlfriend," or "I had a quarrel with my wife," or "We had a family quarrel," or "I was angry with my boss." (Labaw 1980:71)

Interpretive questions: Questions included in a questionnaire or interview schedule to help explain answers to other important questions.

Careers and Research

Grant A. Bacon, Research Associate

Grant Bacon graduated with degrees in history education and political science from the University of Delaware in 1998. He initially aspired to give back to the community, especially by helping young people as a teacher. Although he started out teaching, he found his calling by working more directly with at-risk youth as a court liaison and eventually as a program coordinator for a juvenile drug court/drug diversion program. While working with these drug court programs, Bacon first came into contact with a university-based center for drug and health studies, which was beginning an evaluation of one such program. In 2001, he accepted an offer to become a research associate with the center, where he has continued to work on many different research projects. Two of his most recent projects include research that investigated factors affecting the reentry experiences for inmates returning to the community and another evaluating a parole program.

Bacon is happy to be working in the field on both qualitative and quantitative research. He loves working with people who share a vision of using research findings to help people in a number of ways, and to give back to the world in a meaningful manner. Every day is different. Some days, Bacon and other associates are on the road visiting criminal justice or health-related facilities or are trying to locate specific individual respondents or study participants. Other days, he may be gathering data, doing intensive interviewing, or administering surveys. He thinks the most rewarding part of his job is helping people who have been part of the criminal justice system and giving them a voice.

Bacon has the following advice for students who are interested in research:

> If doing research interests you, ask your teachers how you can gain experience through internships or volunteering. Be sure to network with as many people from as many human services organizations as possible. Being familiar with systems like geographic information systems (GIS) and data analysis is becoming important as well. If you did not receive this training during your undergraduate studies, many community colleges offer introductory and advanced classes in GIS, Microsoft Excel, Access, and SPSS. Take them!

Were these respondents lying in response to the first question? Probably not. More likely, they simply didn't interpret their own behavior in terms of general concepts such as *emotional state.* But their responses to the first question were likely to be misinterpreted without the further detail provided by answers to the second.

Consider five issues when developing interpretive questions—or when you review survey results and need to consider what the answers tell you:

1. *What do the respondents know?* Answers to many questions about current events and government policies are almost uninterpretable without also learning what the respondents know.
2. *What relevant experiences do the respondents have?* Such experiences undoubtedly color the responses. For example, the meaning of opinions about crime and punishment may differ greatly between those who have been crime victims themselves and those who have not.
3. *How consistent are the respondents' attitudes, and do they express some larger perspective or ideology?* An employee who seeks more wages because he or she believes that all employer profits result from exploitation is expressing a different sentiment from one who seeks more wages because he or she really wants a more expensive car with which to impress his or her neighbors.
4. *Are respondents' actions consistent with their expressed attitudes?* We probably should interpret differently the meaning of expressed support for gender equality from married men who help with household chores and those who do not. Questions

about behavior may also provide a better way to assess orientations than will questions about attitudes. Patricia Labaw (1980:100) points out that “the respondent’s actual purchase of life insurance is a more accurate representation of what he believes about his life insurance needs than anything he might say in response to a direct question” about whether it is important to carry life insurance.

5. *How strongly are the attitudes held?* The attitudes of those with stronger beliefs are more likely to be translated into action than are attitudes that are held less strongly. Just knowing the level of popular support for, say, abortion rights or gun control thus fails to capture the likelihood of people to march or petition their representatives on behalf of the cause; we also need to know what proportion of supporters feel strongly (Schuman and Presser 1981:chap. 9). Thus, rather than just asking unmarried respondents if they wish to remarry, Mirowsky and Ross (2001:1) used the following question and response choices to measure strength of desire to remarry in their telephone survey:

How much would you like to get remarried someday?

Would you say . . .

1. <1> Not at all,
2. <2> Somewhat, or
3. <3> Very much?

The qualitative insights produced by open-ended questions (see Chapter 4) can be essential for interpreting the meaning of fixed responses. For example, Renee Anspach (1991) asked administrators, case managers, clients, and family members in four community mental health systems whether their programs were effective. They usually rated their programs as effective when given fixed-choice responses. However, their responses to a series of open-ended questions pointed to many program failings. Anspach concluded that the respondents’ positive answers to her initial question reflect their desire to make the program appear effective, for several reasons: Administrators wanted to maintain funding and employee morale, and case managers wanted to encourage cooperation by clients and their families, as well as to deflect blame for problems to clients, families, or system constraints.

Maintain Consistent Focus

A survey (with the exception of an omnibus survey) should be guided by a clear conception of the research problem under investigation and the population to be sampled. Does the study seek to describe some phenomenon in detail, to explain some behavior, or to explore some type of social relationship? Until the research objective is formulated clearly, survey design cannot begin. Throughout the process of questionnaire design, this objective should be the primary basis for making decisions about what to include and exclude and what to emphasize or treat in a cursory fashion. Moreover, the questionnaire should be viewed as an integrated whole, in which each section and every question serves a clear purpose related to the study's objective and complements other sections or questions.

In the News Research in the News: Social Interaction Critical for Mental and Physical Health

For Further Thought?

Although it was individuals who had been surveyed, when Lisa F. Berkman and S. Leonard Syme (1979) analyzed follow-up data a decade after the 1965 Human Population Laboratory survey of 6,928 adults in Alameda County, California, they found that it was connections between people that made the most difference in their mortality risk—social ties were even more important than socioeconomic status, health practices such as smoking, and use of preventive health services. This conclusion from a survey of the general population is consistent with findings in surveys of patients, randomized trials of interventions, and analyses of insurance records.

1. What strengths and weaknesses of using surveys to study the influences of social ties on health can you suggest?
2. Based on your own experience, what are some of the questions survey researchers should use to operationalize survey respondents' social connections?

News source: Brody, Jane E. 2017. "Friends With Health Benefits." *The New York Times,* June 13, p. D5.

Surveys often include too many irrelevant questions and fail to include questions that, the researchers realize later, are crucial. One way to ensure that possibly relevant questions are asked is to use questions suggested by prior research, theory, experience, or experts (including participants) who are knowledgeable about the setting under investigation. Of course, not even the best researcher can anticipate the relevance of every question. Researchers tend to try to avoid "missing something" by erring on the side of extraneous questions (Labaw 1980:40).

Order the Questions

The order in which questions are presented will influence how respondents react to the questionnaire as a whole and how they may answer some questions (Schwarz 2010:47). As a first step, the individual questions should be sorted into broad thematic categories, which then become separate sections in the questionnaire. For example, the National Longitudinal Survey of Youth 2014 Young Adult questionnaire contained sections on “Regular Schooling,” “Health,” “Attitudes,” “Dating/Relationship History,” and many others (U.S. Bureau of Labor Statistics 2014). Both the sections and the questions within the sections should be organized in a logical order that would make sense in a conversation. Throughout the design process, the grouping of questions in sections and the ordering of questions within sections should be adjusted to maximize the questionnaire’s overall coherence.

The first question deserves special attention, particularly if the questionnaire is to be self-administered. This question signals to the respondent what the survey is about, whether it will be interesting, and how easy it will be to complete. For these reasons, the first question should connect to the primary purpose of the survey, be interesting and easy, and apply to everyone in the sample (Dillman 2000:92–94). Mirowsky and Ross (1999) began their survey about health and related issues with a question about the respondent’s overall health:

First, I’d like to ask you about your health. In general, would you say your health is . . .

1. <1> Very good,
2. <2> Good,
3. <3> Satisfactory,
4. <4> Poor, or
5. <5> Very poor?

One or more filter or screening questions may also appear early in the survey to identify respondents for whom the questionnaire is not intended or perhaps to determine which sections of a multipart questionnaire a respondent is to skip (Peterson 2000:106–107).

Question order can lead to context effects when one or more questions influence how subsequent questions are interpreted (Schober 1999:88–89). For example, when a sample of the general public was asked, “Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?” Fifty-eight percent said yes. However, when this question was preceded by a less permissive question that asked whether the respondent would allow abortion of a defective fetus, only 40% said yes. Asking the question about a defective fetus altered respondents’ frame of reference, perhaps by making abortion simply to avoid having more children seem frivolous by comparison (Turner and Martin 1984:135). Context effects have also been identified in the measurement of general happiness, in what is termed a part–whole question effect (Peterson 2000:113). Married people tend to report that they are happier "in general" if the general happiness question is preceded by a question about their happiness with their marriage (Schuman and Presser 1981:23–77).

Prior questions can influence how questions are comprehended, what beliefs shape responses, and whether comparative judgments are made (Tourangeau 1999). The potential for context effects is greatest when two or more questions concern the same issue or closely related issues, as in the example of the two questions about abortion. The impact of question order also tends to be greater for general, summary-type questions, as with the example about general happiness.

Context effects can be identified empirically if the question order is reversed on a subset of the questionnaires (the so-called split-ballot design) and the results compared. However, knowing that a context effect occurs does not tell us which order is best. Reviewing the overall survey goals and any other surveys with which comparisons should be made can help us decide on question order. What is most important is to be aware of the potential for problems resulting from question order and to evaluate carefully the likelihood of context effects in any particular questionnaire. Those who report survey results should mention, at least in a footnote, the order in which key questions were asked when more than one question about a topic was used (Labaw 1980). An alternative approach is to randomize the order in which key questions are presented, so that any effects of question order cancel each other out.

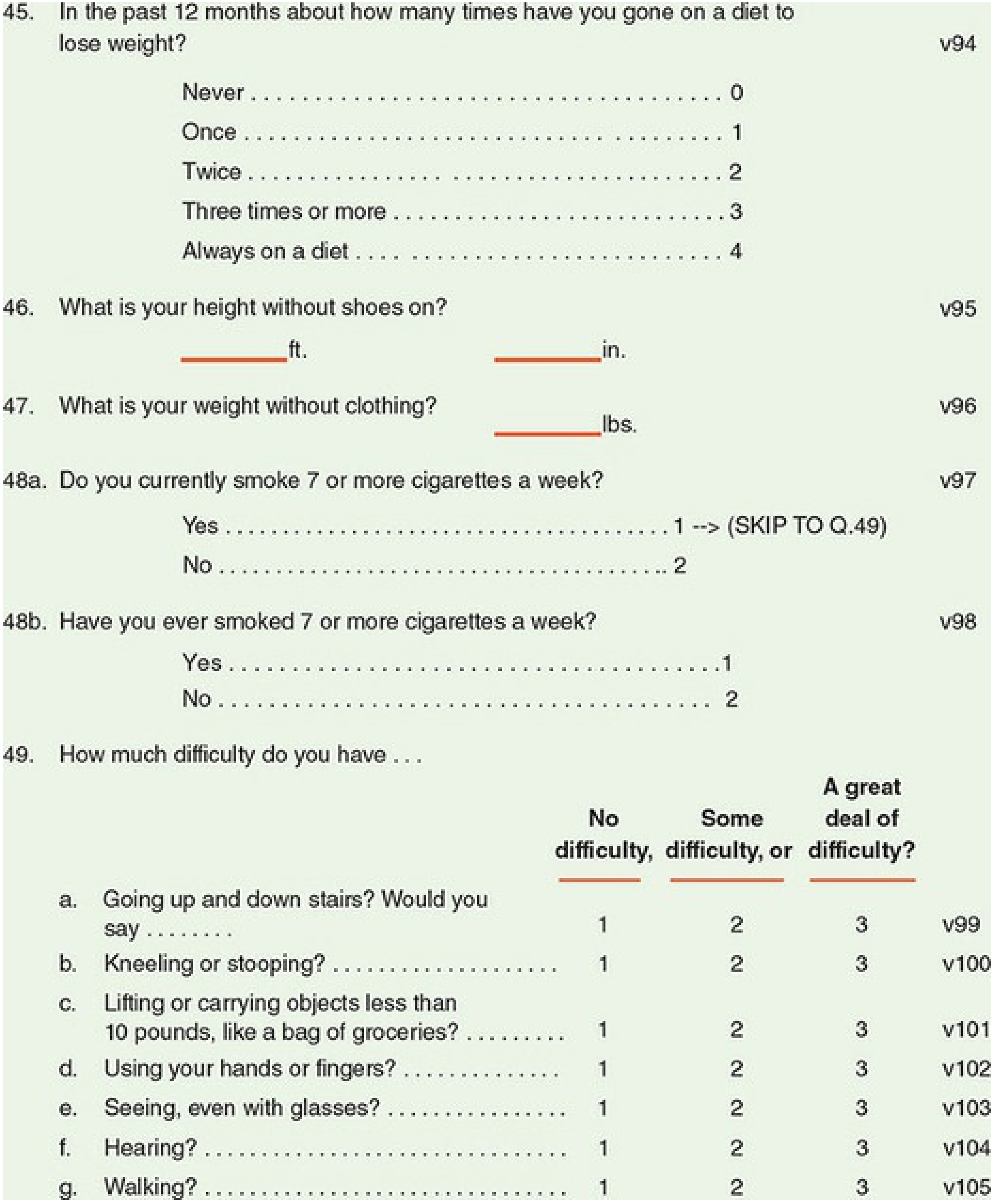
Context effects: When one or more survey questions influence how subsequent questions are interpreted.

Part–whole question effects: When responses to a general or summary survey question about a topic are influenced by responses to an earlier, more specific question about that topic.

Some questions may be presented in a *matrix* format. Matrix questions are a series of questions that concern a common theme and that have the same response choices. The questions are written so that a common initial phrase applies to each one (see Question 49 in Exhibit 8.8). This format shortens the questionnaire by reducing the number of words that must be used for each question. It also emphasizes the common theme among the questions and so invites answering each question in relation to other questions in the matrix. It is very important to provide an explicit instruction to "Check one response on each line" in a matrix question, because some respondents will think that they have completed the entire matrix after they have responded to just a few of the specific questions.

Matrix questions: A series of survey questions that concern a common theme and that have the same response choices.

Exhibit 8.8 A Page From Ross's Interview Schedule



45. In the past 12 months about how many times have you gone on a diet to lose weight? v94

Never 0
Once 1
Twice 2
Three times or more 3
Always on a diet 4

46. What is your height without shoes on? v95

________ft. ________in.

47. What is your weight without clothing? ________lbs. v96

48a. Do you currently smoke 7 or more cigarettes a week? v97

Yes .... 1 --> (SKIP TO Q.49)
No 2

48b. Have you ever smoked 7 or more cigarettes a week? v98

Yes 1
No 2

49. How much difficulty do you have . . .

	No difficulty,	Some difficulty, or	A great deal of difficulty?	
a. Going up and down stairs? Would you say	1	2	3	v99
b. Kneeling or stooping?	1	2	3	v100
c. Lifting or carrying objects less than 10 pounds, like a bag of groceries?	1	2	3	v101
d. Using your hands or fingers?	1	2	3	v102
e. Seeing, even with glasses?	1	2	3	v103
f. Hearing?	1	2	3	v104
g. Walking?	1	2	3	v105

Source: From Catherine E. Ross, Work, Family, and the Sense of Control (1990). Reprinted with permission of the author.

Make the Questionnaire Attractive

An attractive questionnaire is more likely to be completed and less likely to confuse either the respondent or, in an interview, the interviewer. An attractive questionnaire also should increase the likelihood that different respondents interpret the same questions in the same way.

Printing a multipage questionnaire in booklet form usually results in the most attractive and simple-to-use questionnaire. Printing on both sides of folded-over legal-size paper (8½" by 14") is a good approach, although pages can be printed on one side only and stapled in the corner if finances are very tight (Dillman 2000:80–86). An attractive questionnaire does not look cramped; plenty of white space—more between questions than within question components—makes the questionnaire appear easy to complete. Response choices are distinguished clearly and consistently, perhaps by formatting them with light print (while questions are formatted with dark print) and keeping them in the middle of the pages. Response choices are listed vertically rather than horizontally across the page.

The proper path through the questionnaire for each respondent is identified with arrows or other graphics and judicious use of spacing and other aspects of layout. Respondents should not be confused about where to go next after they are told to skip a question. Instructions should help route respondents through skip patterns, and such skip patterns should be used infrequently. Instructions should also explain how each type of question is to be answered (e.g., by circling a number or writing a response) in a neutral way that isn't likely to influence responses. Some distinctive formatting should be used to identify instructions.

The visual design of a questionnaire has more subtle effects on how respondents answer questions. Seemingly minor differences, such as whether responses are grouped under headings or just listed, whether separate response choices are provided or just the instruction to write in a response from a list of choices, and how much space there is between response choices can all affect the distribution of responses to a question (Dillman and Christian 2005:43–48).

Exhibit 8.8 contains portions of the questionnaire Ross (1990) used in a previous phone survey about aging and health. This page illustrates three of the features that I have just reviewed: (1) numeric designation of response choices, (2) clear instructions, and (3) an attractive, open layout. Because this questionnaire was read over the phone, rather than being self-administered, there was no need for more explicit instructions about the matrix question (Question 49) or for a more distinctive format for the response choices (Questions 45 and 48). A questionnaire designed to be self-administered also should include these additional features.

Consider Translation

Should the survey be translated into one or more languages? In the 21st century, no survey plan in the United States or many other countries can be considered complete until this issue has been considered. In the United States in 2014, 13% of persons aged 18 years and older were foreign born (U.S. Census Bureau 2014) and in 2013, one third of Hispanics age 5 or older in the United States are not proficient in English (Krogstad, Lopez, and Rohal 2015). Depending on the specific region or group that is surveyed, these proportions can be much higher and can include persons fluent in various languages (with Spanish being the most common). Although English becomes the primary language spoken by almost all children of immigrants, many first-generation immigrants are not fluent in English (Hakimzadeh and Cohn 2007:i; Pew Hispanic Center 2008:Table 21). As a result, they can be included in a survey only if it is translated into their native language.

When immigrants are a sizable portion of a population, omitting them from a survey can result in a misleading description of the population. Foreign-born persons in the United States tend to be younger than native-born persons and their average income is lower (Pew Hispanic Center 2008:Tables 8a, 29). They also are more likely to be married, to be in a household with five or more family members, and to have less than a high school education (Pew Hispanic Center 2008:Tables 13, 18, 22). However, none of these differences are true for all immigrant groups. In particular, persons from South and East Asia and the Middle East tend to have more education and higher incomes than do persons born in the United States (Pew Hispanic Center 2008:Tables 22, 29).

So, survey researchers find increasingly that they must translate their questionnaires into one or more languages to represent the population of interest. This does not simply mean picking up a bilingual dictionary, clicking “translate” in a web browser, or hiring a translator to translate the questions and response choices word for word. Such a literal translation may not result in statements that are interpreted in the same way to non-English speakers. The U.S. Census Bureau’s (2006) guidelines for translation designate the literal translation as only one step in the process. What is needed is to achieve some equivalence of the concepts in different cultures (Church 2010:154–159). The U.S. Census Bureau and the World Health Organization (n.d.) recommend that questionnaires be translated by a team that includes trained translators, persons who are specialists in the subject matter of the survey, persons with expertise in questionnaire design, and experts with several of these skills who can review the translation and supervise a pretest (Pan and de la Puente 2005).

A properly translated questionnaire will be

- *Reliable:* conveys the intended meaning of the original text
- *Fluent:* reads well and makes sense in the target language
- *Appropriate:* the style, tone, and function are appropriately transferred

Needless to say, this translation process adds cost and complexity to survey design.

Organizing Surveys

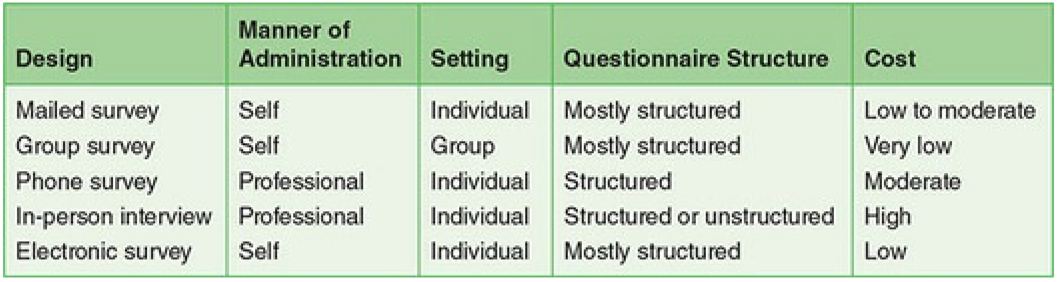
There are five basic social science survey designs: (1) mailed, (2) group-administered, (3) phone, (4) in-person, and (5) electronic. Survey researchers can also combine elements of two or more of these basic designs in mixed-mode surveys. Exhibit 8.9 summarizes the typical features of the five basic survey designs.

Manner of administration.

The five survey designs differ in the manner in which the questionnaire is administered (see Exhibit 8.9). Mailed, group, and electronic surveys are completed by the respondents themselves. During phone and in-person interviews, the researcher or a staff person asks the questions and records the respondent's answers. However, new mixed-mode surveys break down these distinctions. For example, in audio computer-assisted self-interviewing (or audio-CASI), the interviewer gives the respondent a laptop and a headset (Tourangeau 2004:790–791). The respondent reads the questions on the computer screen, hears the questions in the headset, and responds by choosing answers on the computer screen.

Electronic survey: A survey that is sent and answered by computer, either through e-mail or on the web.

Exhibit 8.9 Typical Features of the Five Survey Designs



Design	Manner of Administration	Setting	Questionnaire Structure	Cost
Mailed survey	Self	Individual	Mostly structured	Low to moderate
Group survey	Self	Group	Mostly structured	Very low
Phone survey	Professional	Individual	Structured	Moderate
In-person interview	Professional	Individual	Structured or unstructured	High
Electronic survey	Self	Individual	Mostly structured	Low

Setting.

Most surveys are conducted in settings where only one respondent completes the survey at a time; most mail and electronic questionnaires and phone interviews are intended for completion by only one respondent. The same is usually true of in-person interviews, although sometimes researchers interview several family members at once. A variant of the standard survey is a questionnaire distributed simultaneously to a group of respondents, who complete the survey while the researcher (or assistant) waits. Students in classrooms are typically the group involved, although this type of group distribution also occurs in surveys of employees and members of voluntary groups.

Questionnaire structure.

Survey designs also differ in the extent to which the researcher structures the content and order of questions in advance. Most mailed, group, phone, and electronic surveys are highly structured, fixing in advance the content and order of questions and response choices. Some of these types of surveys, particularly mailed surveys, may include open-ended questions (respondents write in their answers rather than checking off one of several response choices). In-person interviews are often highly structured, but they may include many questions without fixed response choices. Moreover, some interviews may proceed from an interview guide rather than a fixed set of questions. In these relatively unstructured interviews, the interviewer covers the same topics with respondents but varies questions according to the respondent's answers to previous questions. Extra questions are added as needed to clarify or explore answers to the most important questions. Computers make it easy for researchers to use complex branching patterns in questionnaires administered in person, on the phone, or on the web because the computer can present different questions based on responses to prior questions (Tourangeau 2004:789).

Cost.

As mentioned earlier, in-person interviews are the most expensive type of survey. Phone interviews are much less expensive, although costs are rising because of the need to make more calls to reach potential respondents. Surveying by mail is cheaper yet. Electronic surveys can be the least expensive method because there are no interviewer costs, no mailing costs, and, for many designs, almost no costs for data entry. However, extra staff time and programming expertise are required to prepare an electronic questionnaire (Tourangeau, Conrad, and Couper 2012).

Because of their different features, the five designs vary in the types of error to which they are most prone and the situations in which they are most appropriate. The different designs can also be improved in different ways by adding some features of the other designs. This section focuses on the various designs' unique advantages and disadvantages and identifies techniques for reducing error within each design and by combining designs.

Mailed, Self-Administered Surveys

A mailed survey is conducted by mailing a questionnaire to respondents, who then administer the survey themselves. The central concern in a mailed survey is maximizing the response rate. Even an attractive questionnaire full of clear questions will probably be returned by no more than 30% of a sample unless extra steps are taken to increase the rate of response. It's just too much bother for most potential recipients; in the language of social exchange theory, the costs of responding are perceived to be much higher than any anticipated rewards for doing so. Of course, a response rate of 30% is a disaster; even a response rate of 60% represents so much nonresponse error that it is hard to justify using the resulting data. Fortunately, the conscientious use of a systematic survey design method can be expected to lead to an acceptable 70% or higher rate of response to most mailed surveys (Dillman 2000).

Mailed survey: A survey involving a mailed questionnaire to be completed by the respondent.

Sending follow-up mailings to nonrespondents is the single most important requirement for obtaining an adequate response rate to a mailed survey. The follow-up mailings explicitly encourage initial nonrespondents to return a completed questionnaire; implicitly, they convey the importance of the effort. Dillman (2000:155–158, 177–188) has demonstrated the effectiveness of a standard procedure for the mailing process:

- A few days before the questionnaire is to be mailed, send a brief letter to respondents that notifies them of the importance of the survey they are to receive.
- Send the questionnaire with a well-designed, personalized cover letter (see the following description), a self-addressed, stamped return envelope, and, if possible, a token monetary reward. The materials should be inserted in the mail-out envelope so that they will all be pulled out together when the envelope is opened (Dillman 2000:174–175). There should be no chance that the respondent will miss something.
- Send a reminder postcard, thanking respondents and reminding nonrespondents, to all sample members 2 weeks after the initial mailing. The postcard should be friendly in tone and must include a phone number for those people who may not have received the questionnaire. It is important that this postcard be sent before most nonrespondents will have discarded their questionnaire, even though this means the postcard will arrive before all those who might have responded to the first mailing have done so.
- Send a replacement questionnaire with a new cover letter only to nonrespondents, 2 to 4 weeks after the initial questionnaire mailing. This cover letter should be a bit shorter and more insistent than the original cover letter. It should note that the

recipient has not yet responded, and it should stress the survey's importance. Of course, a self-addressed, stamped return envelope must be included.

- The final step is taken 6 to 8 weeks after the initial survey mailing. This step uses a different mode of delivery (either priority or special delivery) or a different survey design—usually an attempt to administer the questionnaire over the phone. These special procedures emphasize the importance of the survey and encourage people to respond.

The cover letter for a mailed questionnaire is critical to the success of a mailed survey. This statement to respondents sets the tone for the questionnaire. A carefully prepared cover letter should increase the response rate and result in more honest and complete answers to the survey questions; a poorly prepared cover letter can have the reverse effects.

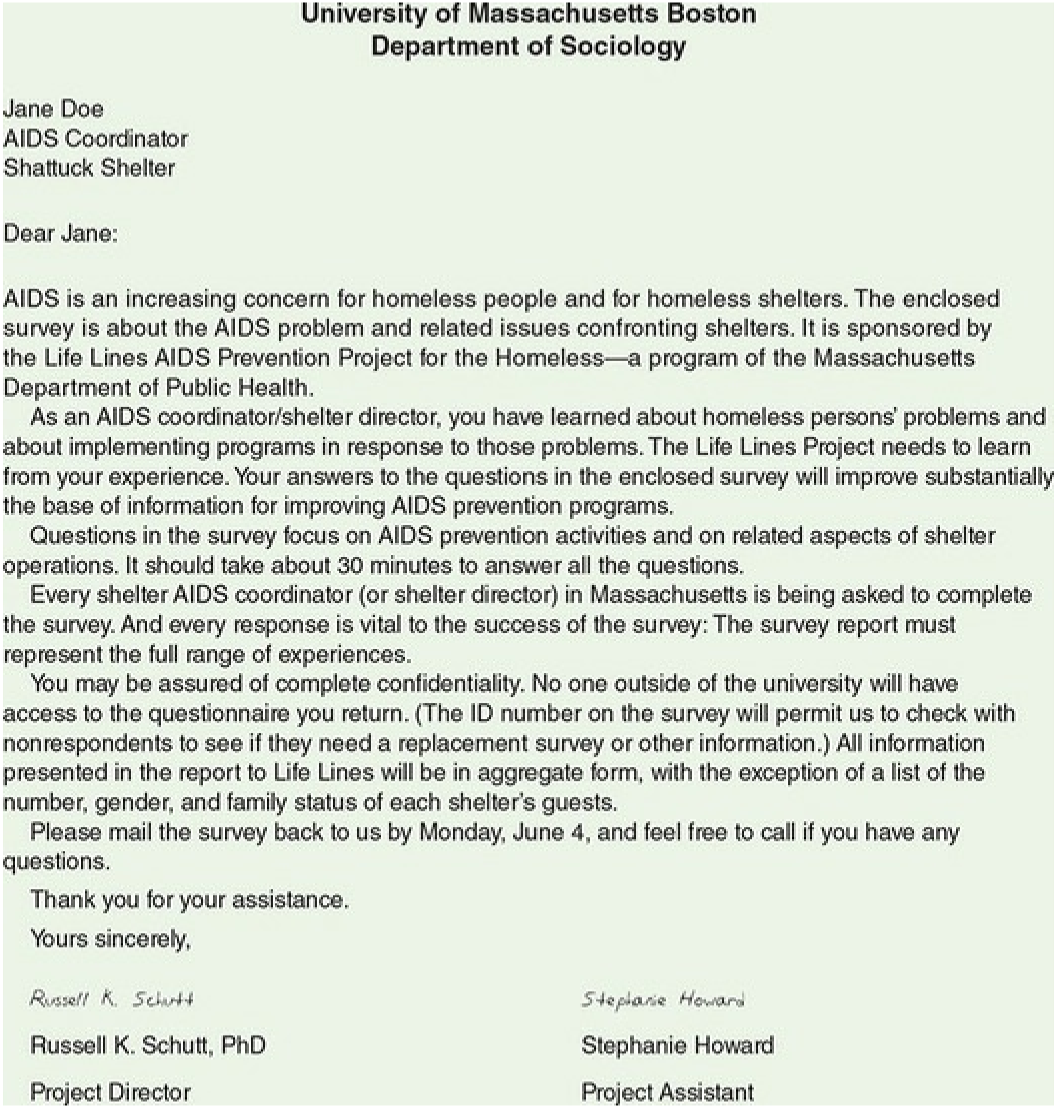
Cover letter: The letter sent with a mailed questionnaire that explains the survey's purpose and auspices and encourages the respondent to participate.

The cover letter or introductory statement must be

- *Credible.* The letter should establish that the research is being conducted by a researcher or organization that the respondent is likely to accept as a credible, unbiased authority. According to one investigation, a sponsor known to respondents may increase the rate of response by as much as 17%. Government sponsors tend to elicit high rates of response. Research conducted by well-known universities and recognized research organizations (e.g., Gallup or RAND) is also usually credible in this sense. The next most credible sponsors are state headquarters of an organization and then other people in a similar field. Publishing firms, students (sorry!), and private associations elicit the lowest response rates.
- *Personalized.* The cover letter should include a personalized salutation (using the respondent's name, e.g., not just "Dear Student"), close with the researcher's signature (blue ballpoint pen is best because that makes it clear that the researcher has personally signed), and refer to the respondent in the second person ("Your participation . . .").
- *Interesting.* The statement should interest the respondent in the contents of the questionnaire. Never make the mistake of assuming that what is of interest to you will also interest your respondents. Try to put yourself in their shoes before composing the statement, and then test your appeal with a variety of potential respondents.
- *Responsible.* Reassure the respondent that the information you obtain will be treated confidentially, and include a phone number to call if the respondent has any questions or would like a summary of the final report. Point out that the respondent's participation is completely voluntary (Dillman 1978:165–172).

Exhibit 8.10 is an example of a cover letter for a questionnaire.

Exhibit 8.10 Sample Questionnaire Cover Letter



University of Massachusetts Boston
Department of Sociology

Jane Doe
AIDS Coordinator
Shattuck Shelter

Dear Jane:

AIDS is an increasing concern for homeless people and for homeless shelters. The enclosed survey is about the AIDS problem and related issues confronting shelters. It is sponsored by the Life Lines AIDS Prevention Project for the Homeless—a program of the Massachusetts Department of Public Health.

As an AIDS coordinator/shelter director, you have learned about homeless persons' problems and about implementing programs in response to those problems. The Life Lines Project needs to learn from your experience. Your answers to the questions in the enclosed survey will improve substantially the base of information for improving AIDS prevention programs.

Questions in the survey focus on AIDS prevention activities and on related aspects of shelter operations. It should take about 30 minutes to answer all the questions.

Every shelter AIDS coordinator (or shelter director) in Massachusetts is being asked to complete the survey. And every response is vital to the success of the survey: The survey report must represent the full range of experiences.

You may be assured of complete confidentiality. No one outside of the university will have access to the questionnaire you return. (The ID number on the survey will permit us to check with nonrespondents to see if they need a replacement survey or other information.) All information presented in the report to Life Lines will be in aggregate form, with the exception of a list of the number, gender, and family status of each shelter's guests.

Please mail the survey back to us by Monday, June 4, and feel free to call if you have any questions.

Thank you for your assistance.

Yours sincerely,

Russell K. Schutt	Stephanie Howard
Russell K. Schutt, PhD	Stephanie Howard
Project Director	Project Assistant

Other steps are necessary to maximize the response rate (Fowler 1988:99–106; Mangione 1995:79–82; Miller and Salkind 2002:164):

- It is particularly important, in self-administered surveys, that the individual questions are clear and understandable to all the respondents because no interviewers will be on hand to clarify the meaning of the questions or to probe for additional details.
- Use no more than a few open-ended questions because respondents are likely to be put off by the idea of having to write out answers.
- Write an ID number on the questionnaire so that you can identify the nonrespondents. This is essential for follow-up efforts. Of course, the ID number must be explained in the cover letter.
- Enclose a token incentive with the survey. A $2 or $5 bill seems to be the best incentive. It is both a reward for the respondent and an indication of your trust that

the respondent will carry out his or her end of the "bargain." The response rate to mailed surveys increases by 19 percentage points, on average, in response to such an incentive (Church 1993). Offering electronic money, a redeemable voucher, or a donation to a charity is less effective than providing cash (Toepoel 2016:109–113). Offering a large monetary reward or some type of lottery ticket only for those who return their questionnaire is actually less effective, apparently because it does not indicate trust in the respondent (Dillman 2000:167–170).

- Include a stamped, self-addressed return envelope with each copy of the questionnaire. This reduces the cost for responding. The stamp helps personalize the exchange and is another indication of trust in the respondent (who could use the stamp for something else). Using a stamp rather than metered postage on the mail-out envelope does not seem to influence the response rate, but it is very important to use first-class rather than bulk-rate postage (Dillman 2000:171–174).
- Consider presurvey publicity efforts. A vigorous advertising campaign increased considerably the response to the 2000 Census mailed questionnaire; the results were particularly successful among minority groups, who had been targeted because of low response rates in the 1990 Census (Holmes 2000).

If Dillman's procedures are followed, and the guidelines for cover letters and questionnaire design also are adhered to, the response rate is almost certain to approach 70%. One review of studies using Dillman's method to survey the general population indicates that the average response to a first mailing will be about 24%; the response rate will rise to 42% after the postcard follow-up, to 50% after the first replacement questionnaire, and to 72% after a second replacement questionnaire is sent by certified mail (Dillman et al. 1974).

The response rate may be higher with particular populations surveyed on topics of interest to them, and it may be lower with surveys of populations that do not have much interest in the topic. When a survey has many nonrespondents, getting some ideas about their characteristics, by comparing late respondents with early respondents, can help determine the likelihood of bias resulting from the low rate of response. If those who returned their questionnaires at an early stage are more educated or more interested in the topic of the questionnaire, the sample may be biased; if the respondents are not more educated or more interested than nonrespondents, the sample will be more credible.

If resources did not permit phone calls to all nonrespondents, a random sample of nonrespondents can be selected and contacted by phone or interviewed in person. It should be possible to secure responses from a substantial majority of these nonrespondents in this way. With appropriate weighting, these new respondents can then be added to the sample of respondents to the initial mailed questionnaire, resulting in a more representative total sample (for more details, see Levy and Lemeshow 1999:398–402).

Related to the threat of nonresponse in mailed surveys is the hazard of incomplete response. Some respondents may skip some questions or just stop answering questions at some point in the questionnaire. Fortunately, this problem does not occur often with well-designed questionnaires. Potential respondents who have decided to participate in the survey usually complete it. But there are many exceptions to this observation because questions that are poorly written, too complex, or about sensitive personal issues simply turn off some respondents. The revision or elimination of such questions during the design phase should minimize the problem. When it does not, it may make sense to impute values for the missing data (in effect, estimate the values of missing data). One imputation procedure would be to substitute the mean (arithmetic average) value of a variable for those cases that have a missing value on the variable (Levy and Lemeshow 1999:404–416).

Group-Administered Surveys

A group-administered survey is completed by individual respondents assembled in a group. The response rate is not usually a major concern in surveys that are distributed and collected in a group setting because most group members will participate. The real difficulty with this method is that it is seldom feasible because it requires what might be called a captive audience. With the exception of students, employees, members of the armed forces, and some institutionalized populations, most populations cannot be sampled in such a setting.

Whoever is responsible for administering the survey to the group must be careful to minimize comments that might bias answers or that could vary between different groups in the same survey (Dillman 2000:253–256). A standard introductory statement should be read to the group that expresses appreciation for their participation, describes the steps of the survey, and emphasizes (in classroom surveys) that the survey is not the same as a test. A cover letter like the one used in mailed surveys also should be distributed with the questionnaires. To emphasize confidentiality, respondents should be given an envelope in which to seal their questionnaires after they are completed.

Another issue of special concern with group-administered surveys is the possibility that respondents will feel coerced to participate and, as a result, will be less likely to answer questions honestly. Also, because administering a survey in this way requires approval of the powers that be—and this sponsorship is made quite obvious by the fact that the survey is conducted on the organization's premises—respondents may infer that the researcher is not at all independent of the sponsor. No complete solution to this problem exists, but it helps to make an introductory statement emphasizing the researcher's independence and giving participants a chance to ask questions about the survey. The sponsor should also understand the need to keep a low profile and to allow the researcher both control over the data and autonomy in report writing. Participation in group-administered surveys of grade school and high school students can be reduced because of the requirement of parental permission, but here the group context can be used to the researcher's advantage. Jane Onoye, Deborah Goebert, and Stephanie Nishimura (2012) at the University of Hawai'i at Manoa found that offering a class a reward such as a pizza if a high rate of participation was achieved led to more parental consent forms being returned than when students were offered a $5 gift card for participating.

Telephone Surveys

In a phone survey, interviewers question respondents over the phone and then record respondents' answers. Phone interviewing became a popular method of conducting surveys in the United States because almost all families had phones by the latter part of the 20th century. But two matters may undermine the validity of a phone survey: not reaching the proper sampling units and not getting enough complete responses to make the results generalizable.

Group-administered survey: A survey that is completed by individual respondents who are assembled in a group.

Phone survey: A survey in which interviewers question respondents over the phone and then record their answers.

Reaching Sample Units

There are three different ways of obtaining a sampling frame of telephone exchanges or numbers: (1) Phone directories provide a useful frame for local studies; (2) a nationwide list of area code or exchange numbers can be obtained from a commercial firm (random digit dialing is used to fill in the last four digits); and (3) commercial firms can provide files based on local directories from around the nation. There are coverage errors with each of these frames: 10% to 15% of directory listings will turn out not to still be valid residential numbers; more than 35% of U.S. households with phones have numbers that are unlisted in directories, and the percentage is as high as 60% in some communities; and less than 25% of the area codes and exchanges in the one national comprehensive list (available from Bell Core Research, Inc.) refer to residential units (Levy and Lemeshow 1999:455–460). In planning a survey, researchers must consider the advantages and disadvantages of these methods for a particular study and develop means to compensate for the weaknesses of the specific method chosen.

Most telephone surveys use random digit dialing at some point in the sampling process (Lavrakas 1987). A machine calls random phone numbers within the designated exchanges, whether or not the numbers are published. When the machine reaches an inappropriate household (such as a business in a survey that is directed to the general population), the phone number is simply replaced with another. Most survey research organizations use special methods (some version of the Mitofsky–Waksberg method) to identify sets of phone numbers that are likely to include working numbers and so make the random digit dialing more efficient (Tourangeau 2004:778–780).

The University of Illinois Survey Research Laboratory used this approach to draw the original sample for Mirowsky and Ross's study of education, social status, and health (Mirowsky 1999). Because the research had a particular focus on health problems related to aging, the researchers used a stratified sampling procedure and oversampled older Americans:

> The survey of Aging, Status, and the Sense of Control (ASOC) is a national telephone probability sample of United States households. A first wave of interviews was completed at the beginning of 1995. Respondents were selected using a prescreened random-digit dialing method that increases the hit rate and decreases standard errors compared with the standard Mitofsky–Waksberg method while producing a sample with the same demographic profile (Lund and Wright 1994; Waksberg 1978). The ASOC survey has two subsamples, designed to produce an 80 percent oversample of persons aged 60 or older. The general sample draws from all households; the oversample draws only from households with one or more seniors. In the general sample the adult (18 or older) with the most recent birthday was selected as respondent. In the oversample the senior (60 or older) with the most recent birthday was selected. For practical reasons the survey was limited to English-speaking adults. Up to 10 callbacks were made to select and contact a respondent, and up to 10 to complete the interview once contact was made. (p. 34)

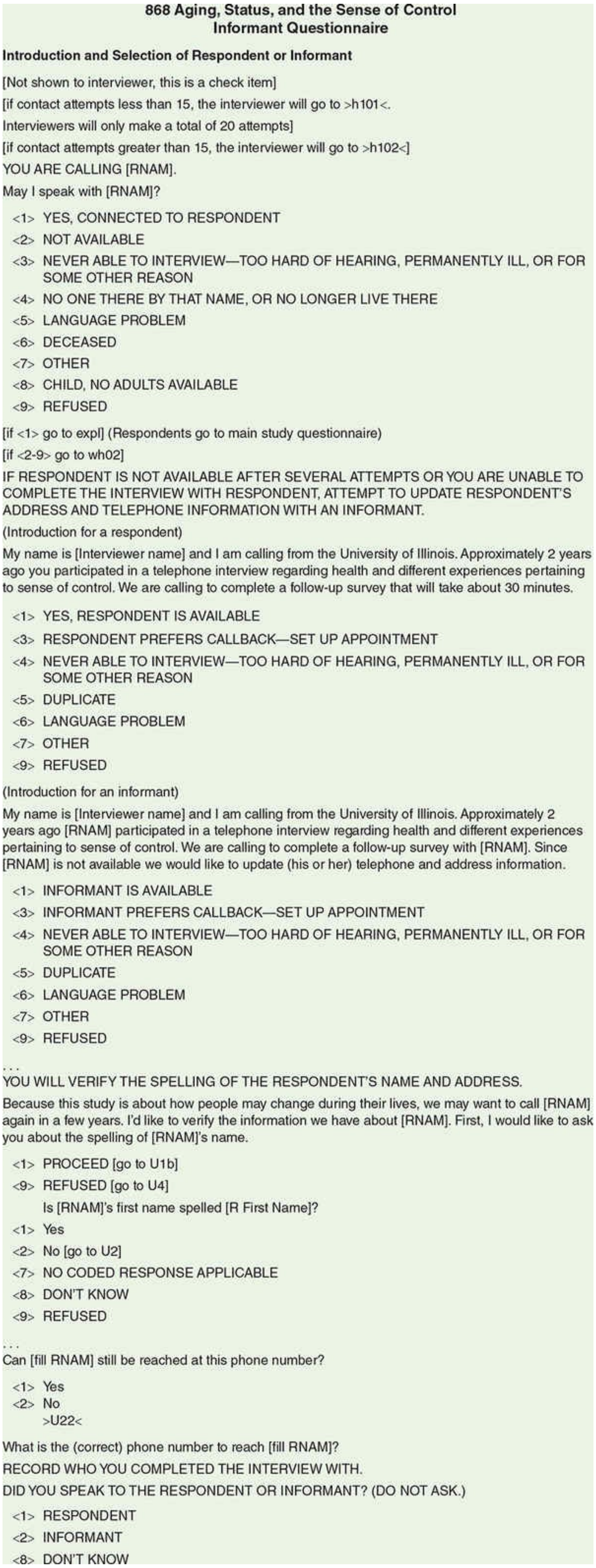
For the third wave of interviews in 2000–2001, SRL planned an intensive effort to contact the original members of the Wave I sample (Mirowsky 1999):

> Attempts will be made to contact and interview all wave 1 respondents, whether or not they were in wave 2, except for individuals who made it clear they did not want to be contacted in the future. A number of new strategies for maximizing follow-up will be tried (Lyberg and Dean 1992; Smith 1995):
>
> (1) Using tested optimal time-of-day/day-of-week callback sequences, lengthening the period of time over which calls are made, and trying a final sequence of five calls three months after an initial sequence of calls fails to make contact; (2) Giving interviewers additional training on establishing rapport and interacting flexibly; (3) Sending advance letters on letterhead to all baseline respondents that include the survey laboratory phone number that will appear on caller ID, an 800 number to call for additional information about the study, several lines of tailored motivational text, and the location of a web page with information about the study, including the e-mail address and phone number of the project coordinator; (4) Sending a letter after first refusal, signed by the

> investigator, explaining the study and the importance of participation, and giving an 800 number to call if they decide to participate; (5) Attempting to find respondents not at the original phone number by using directory assistance, the Equifax database, and six web database programs and search engines; (6) Interviewing persons other than the respondent who answer the phone or persons previously identified by the respondent as likely to know their whereabouts, to locate the respondent or identify a likely reason for noncontact (e.g., passed away, moved to a nursing home, too sick to participate, retired and moved away). (pp. 34–35)

However households are contacted, the interviewers must ask a series of questions at the start of the survey to ensure that they are speaking to the appropriate member of the household. Exhibit 8.11 displays a portion of the instructions that SRL used to select the appropriate respondent for Mirowsky and Ross's phone survey about education and health. This example shows how appropriate and inappropriate households can be distinguished in a phone survey, so that the interviewer is guided to the correct respondent.

Exhibit 8.11 Phone Interview Procedure for Respondent Designation



868 Aging, Status, and the Sense of Control
Informant Questionnaire

Introduction and Selection of Respondent or Informant

[Not shown to interviewer, this is a check item]

[if contact attempts less than 15, the interviewer will go to >h101<.

Interviewers will only make a total of 20 attempts]

[if contact attempts greater than 15, the interviewer will go to >h102<]

YOU ARE CALLING [RNAM].

May I speak with [RNAM]?

<1> YES, CONNECTED TO RESPONDENT
<2> NOT AVAILABLE
<3> NEVER ABLE TO INTERVIEW—TOO HARD OF HEARING, PERMANENTLY ILL, OR FOR SOME OTHER REASON
<4> NO ONE THERE BY THAT NAME, OR NO LONGER LIVE THERE
<5> LANGUAGE PROBLEM
<6> DECEASED
<7> OTHER
<8> CHILD, NO ADULTS AVAILABLE
<9> REFUSED

[if <1> go to expl] (Respondents go to main study questionnaire)

[if <2-9> go to wh02]

IF RESPONDENT IS NOT AVAILABLE AFTER SEVERAL ATTEMPTS OR YOU ARE UNABLE TO COMPLETE THE INTERVIEW WITH RESPONDENT, ATTEMPT TO UPDATE RESPONDENT'S ADDRESS AND TELEPHONE INFORMATION WITH AN INFORMANT.

(Introduction for a respondent)

My name is [Interviewer name] and I am calling from the University of Illinois. Approximately 2 years ago you participated in a telephone interview regarding health and different experiences pertaining to sense of control. We are calling to complete a follow-up survey that will take about 30 minutes.

<1> YES, RESPONDENT IS AVAILABLE
<3> RESPONDENT PREFERS CALLBACK—SET UP APPOINTMENT
<4> NEVER ABLE TO INTERVIEW—TOO HARD OF HEARING, PERMANENTLY ILL, OR FOR SOME OTHER REASON
<5> DUPLICATE
<6> LANGUAGE PROBLEM
<7> OTHER
<9> REFUSED

(Introduction for an informant)

My name is [Interviewer name] and I am calling from the University of Illinois. Approximately 2 years ago [RNAM] participated in a telephone interview regarding health and different experiences pertaining to sense of control. We are calling to complete a follow-up survey with [RNAM]. Since [RNAM] is not available we would like to update (his or her) telephone and address information.

<1> INFORMANT IS AVAILABLE
<3> INFORMANT PREFERS CALLBACK—SET UP APPOINTMENT
<4> NEVER ABLE TO INTERVIEW—TOO HARD OF HEARING, PERMANENTLY ILL, OR FOR SOME OTHER REASON
<5> DUPLICATE
<6> LANGUAGE PROBLEM
<7> OTHER
<9> REFUSED

. . .

YOU WILL VERIFY THE SPELLING OF THE RESPONDENT'S NAME AND ADDRESS.

Because this study is about how people may change during their lives, we may want to call [RNAM] again in a few years. I'd like to verify the information we have about [RNAM]. First, I would like to ask you about the spelling of [RNAM]'s name.

<1> PROCEED [go to U1b]
<9> REFUSED [go to U4]
Is [RNAM]'s first name spelled [R First Name]?
<1> Yes
<2> No [go to U2]
<7> NO CODED RESPONSE APPLICABLE
<8> DON'T KNOW
<9> REFUSED

. . .

Can [fill RNAM] still be reached at this phone number?

<1> Yes
<2> No
>U22<

What is the (correct) phone number to reach [fill RNAM]?

RECORD WHO YOU COMPLETED THE INTERVIEW WITH.

DID YOU SPEAK TO THE RESPONDENT OR INFORMANT? (DO NOT ASK.)

<1> RESPONDENT
<2> INFORMANT
<8> DON'T KNOW

Source: Ross (1990:7).

Maximizing Response to Phone Surveys

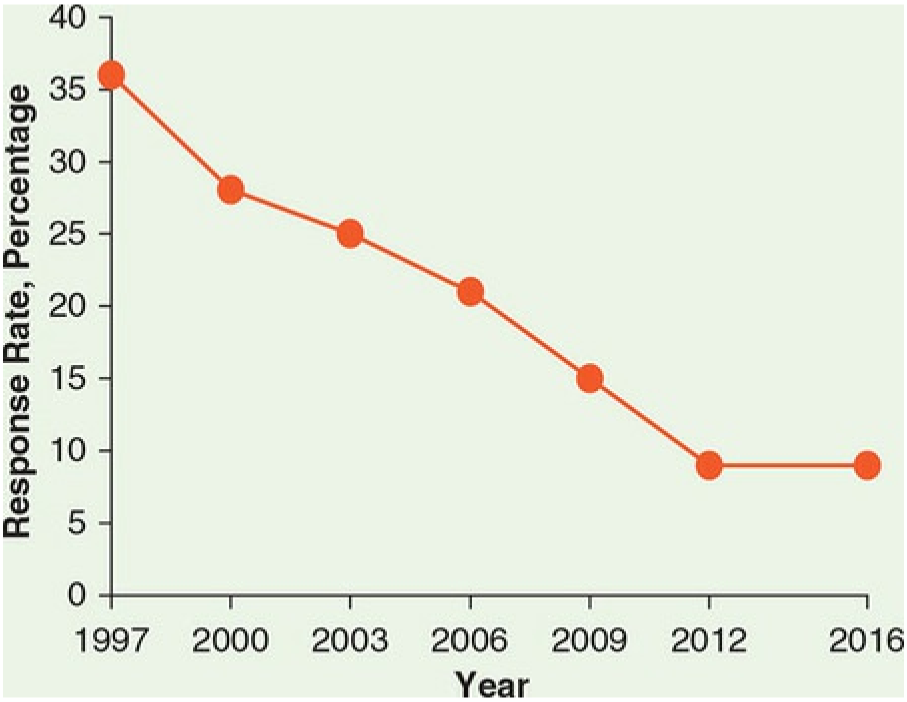
Four issues require special attention in phone surveys. First, because people often are not home, multiple callbacks will be needed for many sample members. Those with more money and education are more likely to be away from home; such persons are more likely to vote Republican, so the results of political polls can be seriously biased if few callback attempts are made (Kohut 1988).

This problem has been compounded in recent years by social changes that are lowering the response rate in phone surveys (Tourangeau 2004:781–783) (see Exhibit 8.12). The Pew Research Center reports a decline in the response rate based on all those sampled from 36% in 1997 to only 9% in 2012 (Kohut et al. 2012).

The number of callbacks needed to reach respondents by telephone has increased greatly in the past 20 years, with increasing numbers of single-person households, dual-earner families, and out-of-home activities. Survey research organizations have increased the usual number of phone contact attempts to 20, from between 4 and 8. The growth of telemarketing has created another problem for telephone survey researchers: Individuals have become more accustomed to "just say no" to calls from unknown individuals and organizations or to simply use their answering machines or caller ID to screen out unwanted calls (Dillman 2000:8, 28). Cell phone users are also harder (and more costly) to contact in phone surveys because their numbers are not in published directories. Households with a cell phone but no landline tend to be younger, so the rate of phone survey participation is declining among those aged 18–34 (Keeter 2008) (see Exhibit 8.13).

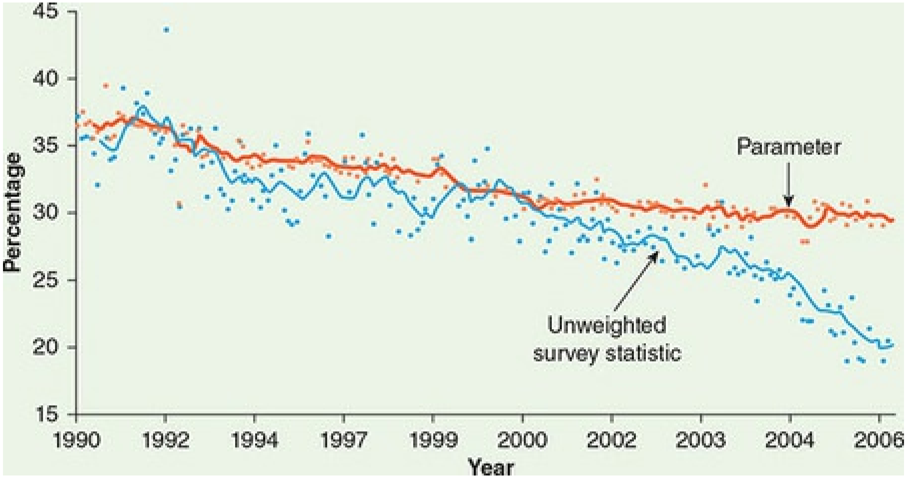
The second issue researchers using phone surveys must cope with are difficulties because of the impersonal nature of phone contact. Visual aids cannot be used, so the interviewer must be able to convey verbally all information about response choices and skip patterns. Instructions to the interviewer must clarify how to ask each question, and response choices must be short. SRL developed the instructions shown in Exhibit 8.14 to clarify procedures for asking and coding a series of questions that Ross (1990) used in another survey to measure symptoms of stress within households.

Exhibit 8.12 Phone Survey Response Rates by Year, 1997–2016



Source: "What Low Response Rates Mean for Telephone Surveys." Pew Research Center, Washington, DC. (May, 2017.) http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/.

Exhibit 8.13 Trend in Percentage, Ages 18 to 34, Responding to Phone Surveys



Source: Pew Research Studies, Keeter (2008).

Third, interviewers must be prepared for distractions because the respondent likely will be interrupted by other household members. Sprinkling interesting questions throughout the questionnaire may help maintain respondent interest. In general, rapport between the interviewer and the respondent is likely to be lower with phone surveys than with in-person interviews, and so respondents may tire and refuse to answer all the questions (Miller and Salkind 2002:317). Distractions are a special problem when respondents are called on a cell phone because they could be driving, in a restaurant or other crowded area, at work, or otherwise involved in activities that make responding difficult and that would not occur in a survey using a landline in the home (AAPOR 2014).

The fourth special consideration for phone surveys is that careful interviewer training is essential. This is how one survey research organization describes its training:

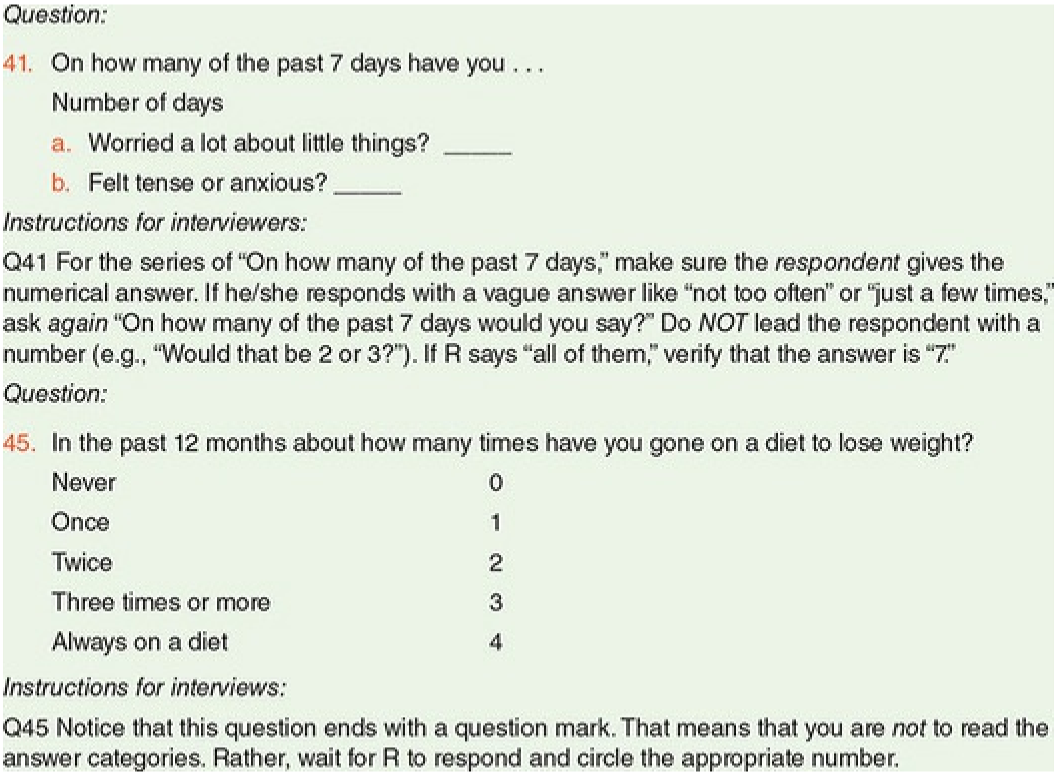
> In preparation for data collection, survey interviewers are required to attend a two-part training session. The first part covers general interviewing procedures and techniques as related to the proposed survey. The second entails in-depth training and practice for the survey. This training includes instructions on relevant subject matter, a question-by-question review of the survey instrument and various forms of role-playing and practice interviewing with supervisors and other interviewers. (J. E. Blair, personal communication to C. E. Ross, April 10, 1989)

Procedures can be standardized more effectively, quality control maintained, and processing speed maximized when phone interviewers use computer-assisted telephone interviews (CATI):

> The interviewing will be conducted using “CATI” (Computer-Assisted Telephone Interviewing). . . . The questionnaire is “programmed” into the computer, along with relevant skip patterns throughout the instrument. Only legal entries are allowed. The system incorporates the tasks of interviewing, data entry, and some data cleaning. (J. E. Blair, personal communication to C. E. Ross, April 10, 1989)

Computer-assisted telephone interview (CATI): A telephone interview in which a questionnaire is programmed into a computer, along with relevant skip patterns, and only valid entries are allowed; incorporates the tasks of interviewing, data entry, and some data cleaning.

Exhibit 8.14 Sample Interviewer Instructions



Question:

41. On how many of the past 7 days have you . . .

    Number of days

    a. Worried a lot about little things? _____

    b. Felt tense or anxious? _____

Instructions for interviewers:

Q41 For the series of "On how many of the past 7 days," make sure the *respondent* gives the numerical answer. If he/she responds with a vague answer like "not too often" or "just a few times," ask *again* "On how many of the past 7 days would you say?" Do *NOT* lead the respondent with a number (e.g., "Would that be 2 or 3?"). If R says "all of them," verify that the answer is "7."

Question:

45. In the past 12 months about how many times have you gone on a diet to lose weight?

Never	0
Once	1
Twice	2
Three times or more	3
Always on a diet	4

Instructions for interviews:

Q45 Notice that this question ends with a question mark. That means that you are *not* to read the answer categories. Rather, wait for R to respond and circle the appropriate number.

Source: Ross (1990).

Computerized interactive voice response (IVR) survey technology allows even greater control over interviewer–respondent interaction. In an IVR survey, respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software. These surveys can also record verbal responses to open-ended questions for later transcription. Although they present some difficulties when many answer choices must be used or skip patterns must be followed, IVR surveys have been used successfully with short questionnaires and when respondents are highly motivated to participate (Dillman 2000:402–411). When these conditions are not met, potential respondents may be put off by the impersonality of this computer-driven approach.

Interactive voice response (IVR): A survey in which respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software.

Phone surveying had for decades been the method of choice for relatively short surveys of the general population. Response rates in phone surveys traditionally tended to be very high —often above 80%—because few individuals would hang up on a polite caller or suddenly stop answering questions (at least within the first 30 minutes or so). Mirowsky and Ross

(2003:207) achieved a response rate of 71.6% for people who could be contacted in their Wave I survey in 1995. However, phone surveying is not a panacea and it should no longer be considered the best method to use for general-purpose surveys. You have already learned of the dramatic decline in phone survey response rates, although this can be mitigated somewhat by extra effort. In a recent phone survey of low-income women in a public health program (Schutt and Fawcett 2005), the University of Massachusetts Center for Survey Research achieved a 55.1% response rate from all eligible sampled clients after a protocol that included as many as 30 contact attempts, although the response rate rose to 72.9% when it was calculated as a percentage of clients who were located (Roman 2005:7). Response rates can be much lower in populations that are young, less educated, and poor. Those who do respond are more likely to be engaged in civic issues than those who do not respond, so estimates of related attitudes and behaviors in phone surveys can be quite biased (Kohut et al. 2012).

In-Person Interviews

What is unique to the in-person interview, compared with the other survey designs, is the face-to-face social interaction between interviewer and respondent. If money is no object, in-person interviewing is often the best survey design.

In-person interviewing has several advantages: Response rates are higher than with any other survey design; questionnaires can be much longer than with mailed or phone surveys; the questionnaire can be complex, with both open-ended and closed-ended questions and frequent branching patterns; the order in which questions are read and answered can be controlled by the interviewer; the physical and social circumstances of the interview can be monitored; and respondents' interpretations of questions can be probed and clarified.

But researchers must be alert to some special hazards resulting from the presence of an interviewer. Respondents should experience the interview process as a personalized interaction with an interviewer who is very interested in the respondent's experiences and opinions. At the same time, however, every respondent should have the same interview experience—asked the same questions in the same way by the same type of person, who reacts similarly to the answers (de Leeuw 2008:318). Therein lies the researcher's challenge—to plan an interview process that will be personal and engaging and yet consistent and nonreactive (and to hire interviewers who can carry out this plan). Careful training and supervision are essential because small differences in intonation or emphasis on particular words can alter respondents' interpretations of questions' meaning (Groves 1989:404–406; Peterson 2000:24). Without a personalized approach, the rate of response will be lower and answers will be less thoughtful—and potentially less valid. Without a consistent approach, information obtained from different respondents will not be comparable—less reliable and less valid.

In-person interview: A survey in which an interviewer questions respondents face-to-face and records their answers.

Balancing Rapport and Control

Adherence to some basic guidelines for interacting with respondents can help interviewers maintain an appropriate balance between personalization and standardization:

- Project a professional image in the interview: that of someone who is sympathetic to the respondent but nonetheless has a job to do.
- Establish rapport at the outset by explaining what the interview is about and how it

will work and by reading the consent form. Ask the respondent if he or she has any questions or concerns, and respond to these honestly and fully. Emphasize that everything the respondent says is confidential.

- During the interview, ask questions from a distance that is close but not intimate. Stay focused on the respondent and make sure that your posture conveys interest. Maintain eye contact, respond with appropriate facial expressions, and speak in a conversational tone of voice.
- Be sure to maintain a consistent approach; deliver each question as written and in the same tone of voice. Listen empathetically, but avoid self-expression or loaded reactions.
- Repeat questions if the respondent is confused. Use nondirective probes—such as “Can you tell me more about that?”—for open-ended questions.

As with phone interviewing, computers can be used to increase control of the in-person interview. In a computer-assisted personal interview (CAPI) project, interviewers carry a laptop computer that is programmed to display the interview questions and to process the responses that the interviewer types in, as well as to check that these responses fall within allowed ranges (Tourangeau 2004:790–791). Interviewers seem to like CAPI, and the data obtained are comparable in quality to data obtained in a noncomputerized interview (Shepherd et al. 1996). A CAPI approach also makes it easier for the researcher to develop skip patterns and experiment with different types of questions for different respondents without increasing the risk of interviewer mistakes (Couper et al. 1998).

The presence of an interviewer may make it more difficult for respondents to give honest answers to questions about socially undesirable behaviors such as drug use, sexual activity, and not voting (Schaeffer and Presser 2003:75). CAPI is valued for this reason because respondents can enter their answers directly in the laptop without the interviewer knowing what their response is. Alternatively, interviewers can simply hand respondents a separate self-administered questionnaire containing the more sensitive questions. After answering these questions, the respondent seals the separate questionnaire in an envelope so that the interviewer does not know the answers. When this approach was used for the GSS questions about sexual activity, about 21% of men and 13% of women who were married or had been married admitted to having cheated on a spouse (“Survey on Adultery” 1993:A20). The degree of rapport becomes a special challenge when survey questions concern issues related to such demographic characteristics as race or gender (Groves 1989). If the interviewer and respondent are similar on the characteristics at issue, the responses to these questions may differ from those that would be given if the interviewer and respondent differ on these characteristics. For example, a white respondent may not disclose feelings of racial prejudice to a black interviewer that he would admit to a white interviewer.

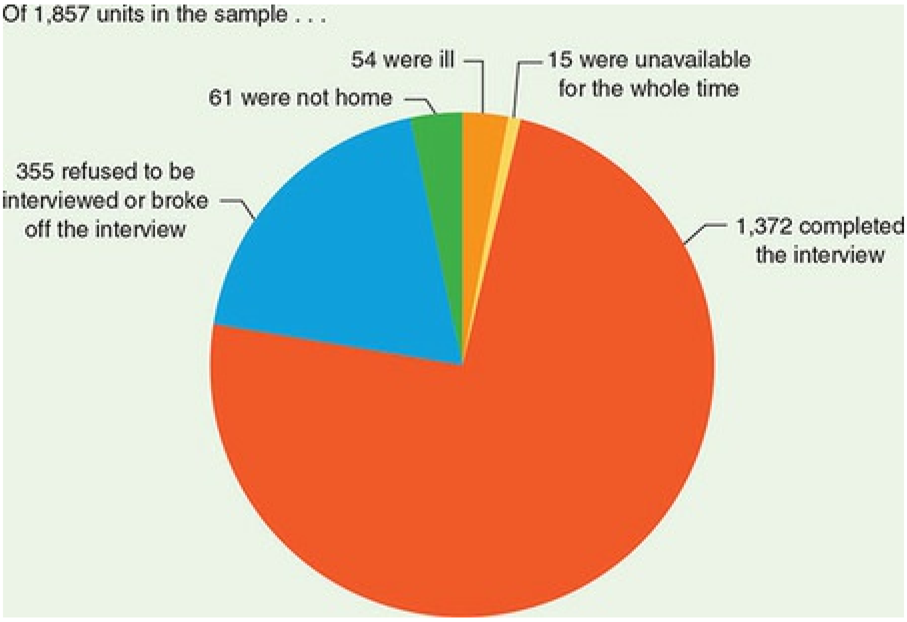
Although in-person interview procedures are typically designed with the expectation that the interview will involve only the interviewer and the respondent, one or more other household members are often within earshot. In a mental health survey in Los Angeles, for example, almost half the interviews were conducted in the presence of another person (Pollner and Adams 1994). It is reasonable to worry that this third-party presence will influence responses about sensitive subjects—even more so because the likelihood of a third party being present may correspond with other subject characteristics. For example, in the Los Angeles survey, another person was present in 36% of the interviews with Anglos, in 47% of the interviews with African Americans, and in 59% of the interviews with Hispanics. However, there is no consistent evidence that respondents change their answers because of the presence of another person. Analysis of this problem with the Los Angeles study found very little difference in reports of mental illness symptoms between respondents who were alone and those who were in the presence of others.

Computer-assisted personal interview (CAPI): A personal interview in which a laptop computer is used to display interview questions and to process responses that the interviewer types in, as well as to check that these responses fall within allowed ranges.

Maximizing Response to Interviews

Even if the right balance has been struck between maintaining control over interviews and achieving good rapport with respondents, in-person interviews still can be problematic. Because of the difficulty of finding all the members of a sample, response rates may suffer. Exhibit 8.15 displays the breakdown of nonrespondents to the 1990 GSS. Of the total original sample of 2,165, only 86% (1,857) were determined to be valid selections of dwelling units with potentially eligible respondents. Among these potentially eligible respondents, the response rate was 74%. The GSS is a well-designed survey using carefully trained and supervised interviewers, so this response rate indicates the difficulty of securing respondents from a sample of the general population even when everything is done "by the book."

Exhibit 8.15 Reasons for Nonresponse in Personal Interviews (1990 General Social Survey)



Source: Data from Davis & Smith 1992:54.

Several factors affect the response rate in interview studies. Contact rates tend to be lower in central cities partly because of difficulties in finding people at home and gaining access to high-rise apartments and partly because of interviewer reluctance to visit some areas at night, when people are more likely to be home (Fowler 1988:45–60). Single-person households also are more difficult to reach, whereas households with young children or elderly adults tend to be easier to contact (Groves and Couper 1998:119–154).

Refusal rates vary with some respondents' characteristics. People with less education participate somewhat less in surveys of political issues (perhaps because they are less aware of current political issues). Less education is also associated with higher rates of "Don't know" responses (Groves 1989). High-income persons tend to participate less in surveys about income and economic behavior (perhaps because they are suspicious about why others want to know about their situation). Unusual strains and disillusionment in a society can also undermine the general credibility of research efforts and the ability of interviewers to achieve an acceptable response rate. These problems can be lessened with an advance letter introducing the survey project and by multiple contact attempts throughout the day and evening, but they cannot be avoided entirely (Fowler 1988:52–53; Groves and Couper 1998). Encouraging interviewers to tailor their response when potential respondents express reservations about participating during the initial conversation can also lead to lower rates of refusal: Making small talk to increase rapport and delaying asking a potential respondent to participate may reduce the likelihood of a refusal after someone first expresses uncertainty about participating (Maynard, Freese, and Schaeffer 2010:810).

Web Surveys

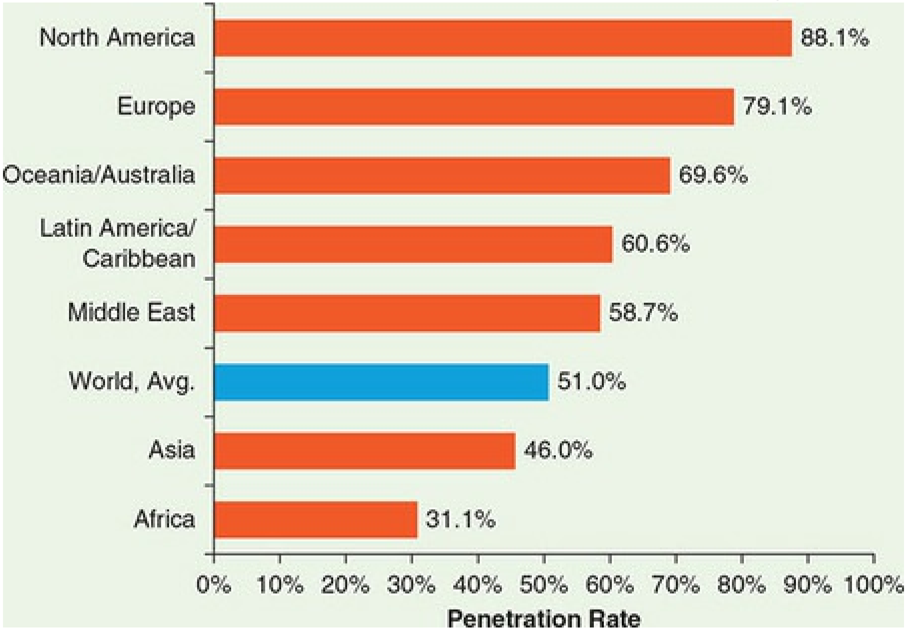
Web surveys have become an increasingly useful survey method for two reasons: growth in the fraction of the population using the Internet and technological advances that make web survey design relatively easy and often superior to printed layouts. Many specific populations have very high rates of Internet use, so a web survey can be a good option for groups such as professionals, middle-class communities, members of organizations, and, of course, college students. Because of the Internet's global reach, web surveys also make it possible to survey large dispersed populations, even in different countries. However, coverage remains a major problem with many populations (Tourangeau et al. 2012). Only 12% of North American households were not connected to the Internet in 2016 (Internet World Stats 2017), so it is becoming possible to survey directly a representative sample of the U.S., Canadian, and Mexican populations on the web—but given a plateau in the rate of Internet connections, the current coverage gap may persist for the near future (Couper and Miller 2008:832). Rates of Internet usage are much lower in other parts of the world, with a worldwide average of 51.0% and rates as low as 31.1% in Africa and 46.0% averaged across all of Asia (see Exhibit 8.16; Internet World Statistics 2017).

The extent to which the population of interest is connected to the web is the most important consideration when deciding whether to conduct a survey through the web. Other considerations that may increase the attractiveness of a web survey include the need for a large sample, for rapid turnaround, or for collecting sensitive information that respondents might be reluctant to discuss in person (Hewson et al. 2016: 41–57; Sue and Ritter 2012:10–11).

Web survey: A survey that is accessed and responded to on the World Wide Web.

There are several different approaches to engaging people in web surveys, each with unique advantages and disadvantages and somewhat different effects on the coverage problem. Many web surveys begin with an e-mail message to potential respondents that contains a direct "hotlink" to the survey website (Gaiser and Schreiner 2009:70). It is important that such e-mail invitations include a catchy phrase in the subject line as well as attractive and clear text in the message itself (Sue and Ritter 2012:110–114). This approach is particularly useful when a defined population with known e-mail addresses is to be surveyed. The researcher can then send e-mail invitations to a representative sample without difficulty. To ensure that the appropriate people respond to a web survey, researchers may require that respondents enter a personal identification number (PIN) to gain access to the web survey (Dillman 2000:378; Sue and Ritter 2012:103–104). Connor, Gray, and Kypri (2010:488) used this approach in their survey of New Zealand undergraduates, ultimately achieving a relatively high 63% response rate:

Exhibit 8.16 Worldwide Internet Penetration Rates by Region, 2017



Source: Copyright 2012, Miniwatts Marketing Group. Reprinted with permission.

> Selected students received a letter which invited them to participate in an internet-based survey as part of the Tertiary Student Health Project, and provided a web address for the survey form. Details of the recruitment and data collection methods have been described in detail previously. Data were collected via a confidential online computerised survey that was completed at a time and place of the respondent's choice.

However, lists of unique e-mail addresses for the members of defined populations generally do not exist outside of organizational settings. Many people have more than one e-mail address, and often there is no apparent link between an e-mail address and the name or location of the person to whom it is assigned. It is not possible to ensure that respondents correctly identify their own characteristics (Toepoel 2016:49). As a result, there is no available method for drawing a random sample of e-mail addresses for people from any general population, even if the focus is only on those with Internet access (Dillman 2007:449). Instead, web survey participants should be recruited from mailings or phone calls to their home addresses, with the web survey link sent to them after they have agreed to participate (Toepoel 2016:72–73). Individuals or families that are willing to participate but do not have Internet access should be provided with a computer (if needed) and an Internet connection. This approach increases the cost of the survey considerably, but it can be used as part of creating the panel of respondents who agree to be contacted for multiple surveys over time. The start-up costs can then be spread across many surveys. Gfk Knowledge Networks is a company that received funding from the U.S. National Science Foundation to create such a web survey panel. CentERdata in the Netherlands also uses this panel approach (Couper and Miller 2008:832–833).

Coverage bias can also be a problem with web surveys that are designed for a population with high levels of Internet use—if the topic of the survey leads some people to be more likely to respond on the web. William Wells, Michael Cavanaugh, Jeffrey Bouffard, and Matt Nobles (2012:461) identified this problem in a comparison of attitudes of students responding to a web survey about gun violence with students at the same university who responded to the same survey administered in classes. Here is their e-mail survey introduction to potential respondents:

> Recently, in response to shootings on university campuses like Virginia Tech and Northern Illinois University, several state legislatures (South Dakota, Texas, Washington) have begun debating whether to change rules banning students and employees from carrying concealed weapons on campus. This is an important public safety issue and the faculty in . . . are interested in knowing how people on this campus feel about it.

Students who responded to the web survey were much more likely to support the right to carry concealed weapons on campus than were those who responded in the classroom survey. In general, having a more extreme attitude motivated people to participate.

When web surveys are administered to “volunteer samples” that are recruited on a particular web site or through social media, the result can be a very large but very biased sample of the larger population (Couper 2000:486–487; Dillman 2000:355). For example, the volunteer sample that participated in the National Geographic Society’s global “Survey 2000” was very large (50,000 respondents), but it represented the type of people who visit the Society’s website (middle class, young, North American) (Witte, Amoroso, and Howard 2000). When Dutch researchers recruited participants for a study on leisure by placing ads on their Facebook pages, the ads were viewed 2.6 million times but “clicked” on only 551 times and led to only 13 persons enrolling in the panel after answering some screening questions (Toepoel 2016:68–69).

Bias in volunteer samples can be reduced somewhat by requiring participants to meet certain inclusion criteria (Selm and Jankowski 2006:440). Another approach to reducing coverage bias with volunteer samples recruited on the Internet is to weight respondents based on key characteristics so that the resulting sample is more comparable to the general population in terms of such demographics as gender, race, age, and education (Couper and Miller 2008:832–833). It appears that weighting can reduce coverage bias by 30% to 60%

(Tourangeau et al. 2012).

The historic Harris Poll (2017) describes such a volunteer panel with adjustments:

> Respondents for this survey were selected from among those who have agreed to participate in Harris Poll surveys. The data have been weighted to reflect the composition of the adult population. Because the sample is based on those who agreed to participate in our panel, no estimates of theoretical sampling error can be calculated.

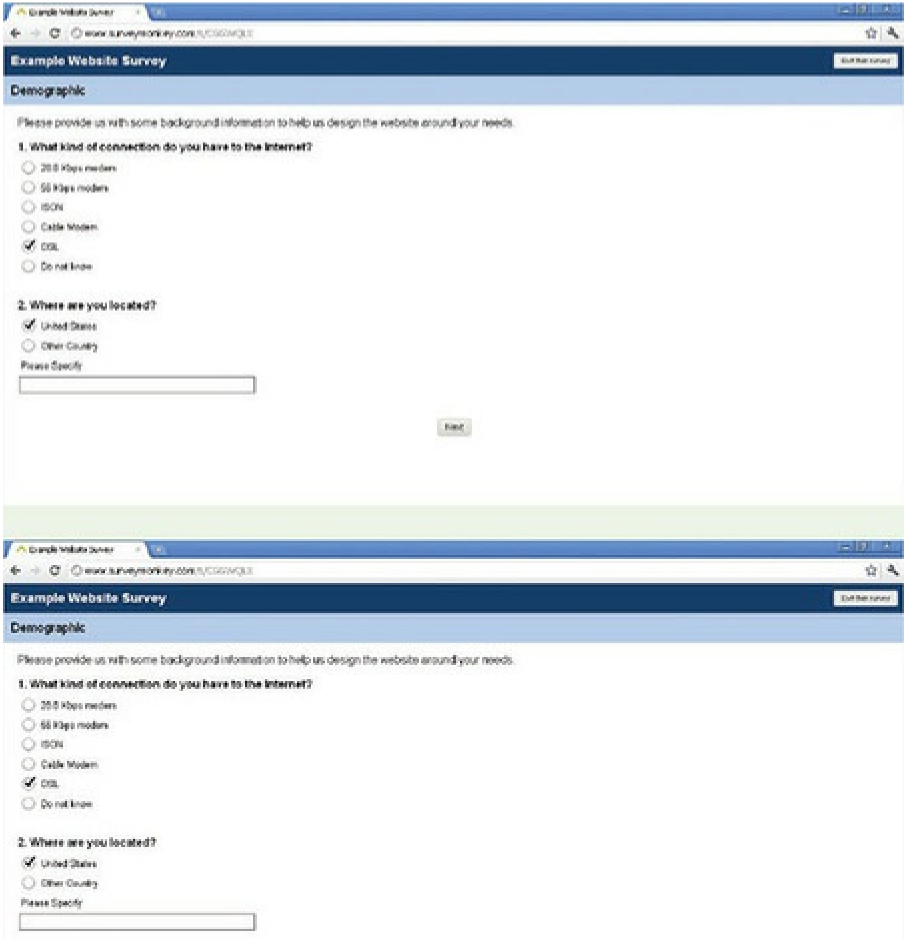
Of course, coverage bias is not as important when a convenience sample will suffice for an exploratory survey about some topic. Audrey Freshman (2012:41) used a web survey of a convenience sample to study symptoms of posttraumatic stress disorder (PTSD) among victims of the Bernie Madoff financial scandal.

> This convenience, nonprobability sample was solicited via direct link to the study placed in online Madoff survivor support groups and comment sections of newspapers and blogs dealing with the event. The study announcement encouraged victims to forward the link to other former investors who might be interested in responding to the survey, thereby creating a snowball effect. The link led directly to a study description and enabled respondents to give informed consent prior to study participation. Participants were assured of anonymity of their responses and were instructed how to proceed in the event of increased feelings of distress as a result of study material. The survey was presumed to take approximately five to 10 minutes to complete. (p. 41)

Although a majority of respondents met clinical criteria for a diagnosis of PTSD, there is no way to know if this sample represents the larger population of Madoff's victims.

In contrast to problems of coverage, web surveys have some unique advantages for increasing measurement validity (Selm and Jankowski 2006; Tourangeau et al. 2012). Questionnaires completed on the web can elicit more honest reports about socially undesirable behavior or experiences, including illicit behavior and victimization in the general population and failing course grades among college students, when compared with results with phone interviews (Kreuter, Presser, and Tourangeau 2008; Parks, Pardi, and Bradizza 2006). Onoye and colleagues (2012) found that conducting a survey on the web increased self-reports of substance use compared with a paper-and-pencil survey. Although they should be short—with about 25–30 questions requiring no more than 10 minutes to complete—web surveys are relatively easy to complete because respondents simply click on response boxes. The survey can be programmed to move respondents easily through sets of questions—not even displaying questions that do not apply to the respondent, thus leading to higher rates of item completion (Hewson et al. 2016:142; Kreuter et al. 2008; Toepoel 2016:35).

Exhibit 8.17 Survey Monkey Web Survey Example



Example Website Survey

Demographic

Please provide us with some background information to help us design the website around your needs.

1. What kind of connection do you have to the Internet?

○ 28.8 Kbps modem
○ 56 Kbps modem
○ ISDN
○ Cable Modem
☑ DSL
○ Do not know

2. Where are you located?

☑ United States
○ Other Country

Please Specify

Next

Example Website Survey

Demographic

Please provide us with some background information to help us design the website around your needs.

1. What kind of connection do you have to the Internet?

○ 28.8 Kbps modem
○ 56 Kbps modem
○ ISDN
○ Cable Modem
☑ DSL
○ Do not know

2. Where are you located?

☑ United States
○ Other Country

Please Specify

Source: Survey Monkey.

Use of the visual, interactive web medium can also help. Pictures, sounds, and animation can be used as a focus of particular questions and graphic and typographic variation can be used to enhance visual survey appeal (see Exhibit 8.17). Definitions of terms can also “pop up” when respondents scroll over them (Dillman 2007:458–459). In these ways, a skilled web programmer can generate a survey layout with many attractive features that make it more likely that respondents will give their answers—and have a clear understanding of the question (Smyth et al. 2004:4–5). Responses can quickly be checked to make sure they fall within the allowable range. Because answers are recorded directly in the researcher’s database, data entry errors are almost eliminated and results can be reported quickly. By taking advantage of these features, Titus Schleyer and Jane Forrest (2000:420) achieved a 74% response rate in a survey of dental professionals who were already Internet users.

SurveyMonkey (www.surveymonkey.com*)* and Qualtrics (www.qualtrics.com) are popular options for designing questionnaires and administering surveys online. They maintain a panel of respondents for researchers who wish to administer a survey to this "SurveyMonkey Audience" (SurveyMonkey n.d.):

> We recruit US survey respondents from the 30+ million people who complete SurveyMonkey surveys each month. They volunteer to join our panel. For every survey our US panelists take, we donate $0.50 to their preferred charity. This attracts people who value giving back and encourages thoughtful, honest participation. (apx. $1 per respondent for 10 questions and 1000 respondents) [free for just 10 questions and 100 respondents].

Despite some clear advantages of some types of web surveys, researchers who use this method must be aware of some important disadvantages. Coverage bias is the single biggest problem with web surveys of the general population and of segments of the population without a high level of Internet access, and none of the different web survey methods fully overcome this problem. Weighting web survey panels of Internet users by demographic and other characteristics does not by itself result in similar responses on many questions with those that are obtained from a mailed survey to a sample of the larger population (Rookey, Hanway, and Dillman 2008). Although providing Internet access to all who agree to participate in a web survey panel reduces coverage bias, many potential respondents do not agree to participate in such surveys: The rate of agreement to participate was 57% in one Knowledge Networks survey and just 41.5% in a survey of students at the University of Michigan (Couper 2000:485–489). Only about one third of Internet users contacted in phone surveys agree to provide an e-mail address for a web survey and then only one third of those actually complete the survey (Couper 2000:488). Web surveys that take more than 15 minutes are too long for most respondents (de Leeuw 2008:322). Some researchers have found that when people are sent a mailed survey that also provides a link to a web survey alternative, they overwhelmingly choose the paper survey (Couper 2000:488). Surveys by phone continue to elicit higher rates of response (Kreuter et al. 2008).

Visual and other highlights that are possible in web surveys should be used with caution to avoid unintended effects on interpretation of questions and response choices (Tourangeau et al. 2012). For example, respondents tend to believe that a response in the middle is the typical response, that responses near each other are related, and that things that look alike are similar. Even minor visual cues can make a difference in responses. In one survey, 5% of respondents shifted their response when one response was given more space relative to others.

Surveys are also now being conducted through social media such as Facebook, on smartphones, and via text messages (Sue and Ritter 2012:119–122). Research continues into the ways that the design of web surveys can influence rates of initial response, the likelihood of completing the survey, and the validity of the responses (Couper, Traugott, and Lamias 2001; Kreuter et al. 2008; Porter and Whitcomb 2003; Tourangeau et al. 2012). At this point, there is reason enough to consider the option of a web survey for many investigations, but proceed with caution and consider carefully their strengths and weaknesses when designing a web survey of any type and when analyzing findings from it.

Mixed-Mode Surveys

Survey researchers increasingly are combining different survey designs to improve the overall participation rate and to take advantage of the unique strengths of different methods. Mixed-mode surveys allow the strengths of one survey design to compensate for the weaknesses of another, and they can maximize the likelihood of securing data from different types of respondents (Dillman 2007:451–453; Selm and Jankowski 2006). For example, a survey may be sent electronically to sample members who have e-mail addresses and mailed to those who don't. Phone reminders may be used to encourage responses to web or paper surveys, or a letter of introduction may be sent in advance of calls in a phone survey (Guterbock 2008). Alternatively, nonrespondents in a mailed survey may be interviewed in person or over the phone. In one comparative study, the response rate to a telephone survey rose from 43% to 80% when it was followed by a mailed questionnaire (Dillman 2007:456). Kristen Olson, Jolene Smyth, and Heather Wood (2012) at the University of Nebraska–Lincoln found that providing a survey in the mode that potential respondents preferred—phone, mailed, or web—increased the overall rate of participation by a small amount. As noted previously, an interviewer may also mix modes by using a self-administered questionnaire to present sensitive questions to a respondent in an in-person interview.

Mixed-mode survey: A survey that is conducted by more than one method, allowing the strengths of one survey design to compensate for the weaknesses of another and maximizing the likelihood of securing data from different types of respondents; for example, nonrespondents in a mailed survey may be interviewed in person or over the phone.

The mixed-mode approach is not a perfect solution. Rebecca Medway and Jenna Fulton (2012) reviewed surveys that gave the option of responding to either a mailed questionnaire or a web questionnaire and found that this reduced the response rate compared with using only a mailed questionnaire. Perhaps the need to choose between the modes or the delay in deciding to start the web survey led some potential respondents not to bother. Respondents to the same question may give different answers because of the survey mode, rather than because they actually have different opinions (Toepoel 2016:98–99). For example, when equivalent samples were asked by phone or mail, "Is the gasoline shortage real or artificial?" many more phone respondents than mail respondents answered that it was "very real" (Peterson 2000:24). Apparently, respondents to phone survey questions tend to endorse more extreme responses to scalar questions (which range from more to less) than do respondents to mail or web surveys (Dillman 2007:456–457). Responses may also differ between questions—one third of the questions in one survey—when asked in web and phone survey modes, even with comparable samples (Rookey et al. 2008:974). When responses differ by survey mode, there is often no way to know which responses are more accurate, although it appears that web surveys are likely to result in more admissions of socially undesirable experiences (Kreuter et al. 2008; Peterson 2000:24). Use of the same question structures, response choices, and skip instructions across modes substantially reduces the likelihood of mode effects, as does using a small number of response choices for each question (Dillman 2000:232–240; Dillman and Christian 2005). Web survey researchers are also identifying considerable effects of visual appearance on the response to questions (Matejka et al. 2016).

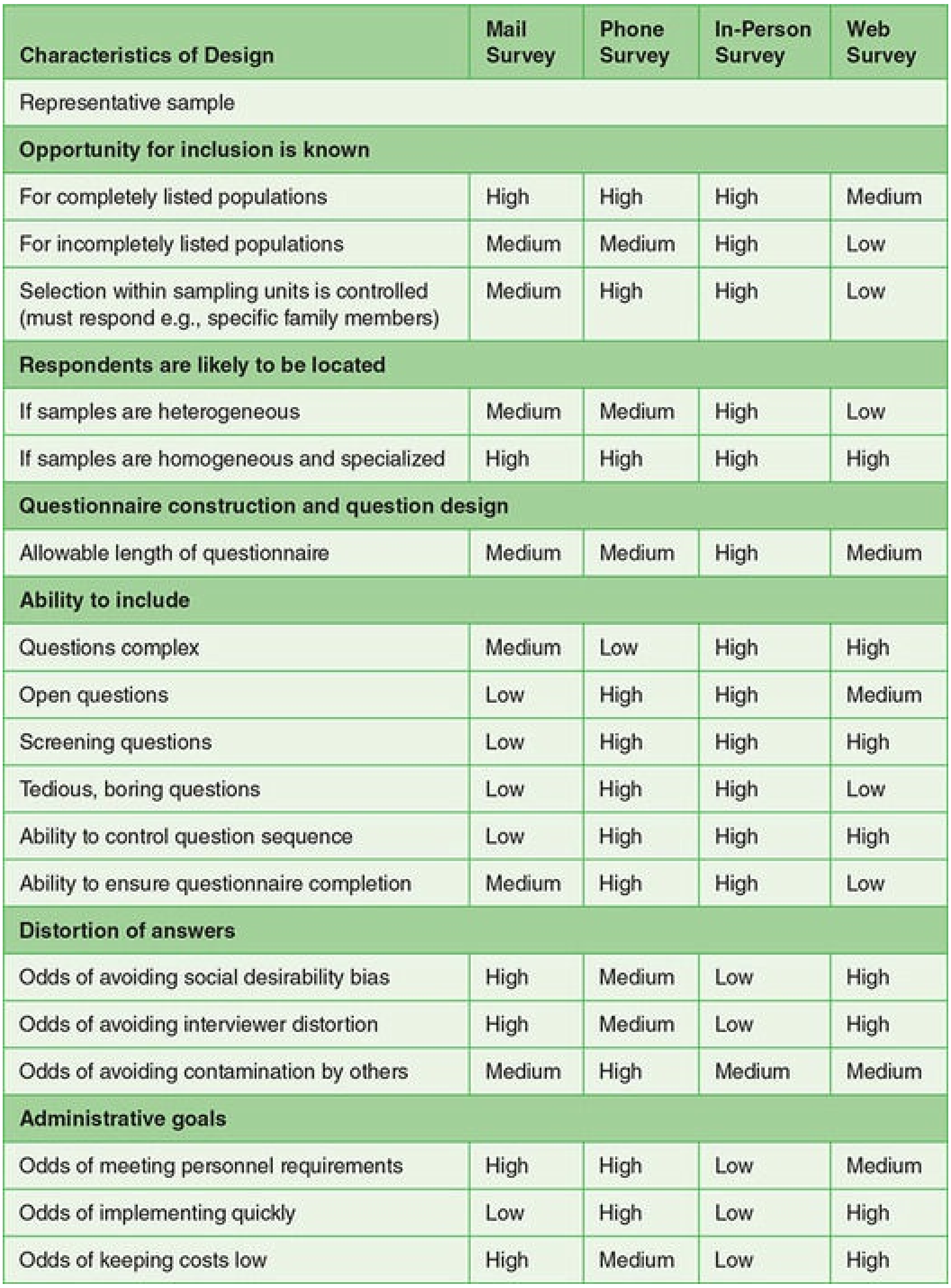
A Comparison of Survey Designs

Which survey design should be used when? Group-administered surveys are similar, in most respects, to mailed surveys, except that they require the unusual circumstance of having access to the sample in a group setting. We therefore don't need to consider this survey design by itself; what applies to mailed surveys applies to group-administered survey designs, with the exception of sampling issues. The features of mixed-mode surveys depend on the survey types that are being combined. Thus, we can focus our comparison on the four survey designs that involve the use of a questionnaire with individuals sampled from a larger population: (1) mailed surveys, (2) phone surveys, (3) in-person surveys, and (4) electronic surveys. Exhibit 8.18 summarizes their strong and weak points.

The most important consideration in comparing the advantages and disadvantages of the four methods is the likely response rate they will generate. Mailed surveys must be considered the least preferred survey design from a sampling standpoint, although declining rates of response to phone surveys are changing this comparison.

Contracting with an established survey research organization for a phone survey is often the best alternative to a mailed survey. The persistent follow-up attempts that are necessary to secure an adequate response rate are much easier over the phone than in person. But, as explained earlier, the process requires an increasing number of callbacks to many households and rates of response have been declining. Current federal law prohibits automated dialing of cell phone numbers, so it is very costly to include the growing number of cell phone–only individuals in a phone survey.

Exhibit 8.18 Advantages and Disadvantages of the Four Survey Designs



Characteristics of Design	Mail Survey	Phone Survey	In-Person Survey	Web Survey
Representative sample				
Opportunity for inclusion is known				
For completely listed populations	High	High	High	Medium
For incompletely listed populations	Medium	Medium	High	Low
Selection within sampling units is controlled (must respond e.g., specific family members)	Medium	High	High	Low
Respondents are likely to be located				
If samples are heterogeneous	Medium	Medium	High	Low
If samples are homogeneous and specialized	High	High	High	High
Questionnaire construction and question design				
Allowable length of questionnaire	Medium	Medium	High	Medium
Ability to include				
Questions complex	Medium	Low	High	High
Open questions	Low	High	High	Medium
Screening questions	Low	High	High	High
Tedious, boring questions	Low	High	High	Low
Ability to control question sequence	Low	High	High	High
Ability to ensure questionnaire completion	Medium	High	High	Low
Distortion of answers				
Odds of avoiding social desirability bias	High	Medium	Low	High
Odds of avoiding interviewer distortion	High	Medium	Low	High
Odds of avoiding contamination by others	Medium	High	Medium	Medium
Administrative goals				
Odds of meeting personnel requirements	High	High	Low	Medium
Odds of implementing quickly	Low	High	Low	High
Odds of keeping costs low	High	Medium	Low	High

Source: Adapted from Dillman 1978: 74–75. Mail and Telephone Surveys: The Total Design Method. Reprinted by permission of John Wiley & Sons, Inc.

In-person surveys are preferable in the possible length and complexity of the questionnaire itself, as well as with respect to the researcher’s ability to monitor conditions while the questionnaire is completed. Mailed surveys often are preferable for asking sensitive questions, although this problem can be lessened in an interview by giving respondents a separate sheet to fill out or a laptop on which to enter their answers. Although interviewers may themselves distort results, either by changing the wording of questions or by failing to record answers properly, survey research organizations can reduce this risk through careful interviewer training and monitoring. Some survey supervisors will have interviews recorded so that they can review the dialogue between interviewers and respondents and provide feedback to the interviewers to help improve their performance. Some survey organizations have also switched to having in-person interviews completed entirely by the respondents on a laptop as they listen to prerecorded questions.

A phone survey limits the length and complexity of the questionnaire but offers the possibility of very carefully monitoring interviewers (Dillman 1978; Fowler 1988:61–73):

> Supervisors in [one organization's] Telephone Centers work closely with the interviewers, monitor their work, and maintain records of their performance in relation to the time schedule, the quality of their work, and help detect and correct any mistakes in completed interviews prior to data reduction and processing. (J. E. Blair, personal communication to C. E. Ross, April 10, 1989)

People interviewed by phone tend to be less interested in the survey than are those interviewed in person, so they tend to satisfice more—apparently in a desire to complete the survey more quickly—and they tend to be less trusting of the survey motives (Holbrook et al. 2003).

The advantages and disadvantages of electronic surveys must be weighed in light of the population that is to be surveyed and capabilities at the time that the survey is to be conducted. At this time, too many people lack Internet connections for survey researchers to use the Internet to survey the general population across the globe, or to avoid a bias toward more educated, younger people even in the United States.

These various points about the different survey designs lead to two general conclusions. First, in-person interviews are the strongest design and generally preferable when sufficient resources and a trained interview staff are available; telephone surveys have many of the advantages of in-person interviews at much less cost, but response rates are an increasing problem. Second, the "best" survey design for any particular study will be determined by the study's unique features and goals rather than by any absolute standard of what the best survey design is.

Ethical Issues in Survey Research

Survey research usually poses fewer ethical dilemmas than do experimental or field research designs. Potential respondents to a survey can easily decline to participate, and a cover letter or introductory statement that identifies the sponsors of, and motivations for, the survey gives them the information required to make this decision. The methods of data collection are quite obvious in a survey, so little is concealed from the respondents. Only in group-administered surveys might the respondents be, in effect, a captive audience (probably of students or employees), and so these designs require special attention to ensure that participation is truly voluntary. (Those who do not wish to participate may be told they can just hand in a blank form.)

The new revised proposed federal regulations to protect human subjects allow most survey research to be exempted from formal review (*Federal Register* 2017:7261, 7262, 7264). Surveys fall within the exemption criteria that you learned about in Chapter 3, which stipulate that research is exempt from review if respondents cannot readily be identified (or procedures are in place to protect their privacy and maintain confidentiality of the data) or if disclosure of their responses would not place them at risk in terms of legal action, financial standing, employability, educational advancement, or reputation.

Confidentiality is most often the primary focus of ethical concern in survey research. Many surveys include essential questions that might, in some way, prove damaging to the subjects if their answers were disclosed. To prevent any possibility of harm to subjects because of the disclosure of such information, the researcher must preserve subject confidentiality. Nobody but research personnel should have access to information that could be used to link respondents to their responses, and even that access should be limited to what is necessary for specific research purposes. Respondents should be identified on the questionnaires only with numbers, and the names that correspond to these numbers should be kept in a safe location—unavailable to staff and others who might otherwise come across them. Follow-up mailings or contact attempts that require linking the ID numbers with names and addresses should be carried out by trustworthy assistants under close supervision. For electronic surveys, encryption technology should be used to make information provided over the Internet secure from unauthorized persons.

Not many surveys can provide true anonymity, so that no identifying information is ever recorded to link respondents with their responses. The main problem with anonymous surveys is that they preclude follow-up attempts to encourage participation by initial nonrespondents, and they prevent panel designs, which measure change through repeated surveys of the same individuals. In-person surveys rarely can be anonymous because an interviewer must, in almost all cases, know the name and address of the interviewee. However, phone surveys that are meant only to sample opinion at one point in time, as in political polls, can safely be completely anonymous. When no future follow-up is desired, group-administered surveys also can be anonymous. To provide anonymity in a mail survey, the researcher should omit identifying codes from the questionnaire but could include a self-addressed, stamped postcard so that the respondent can notify the researcher that the questionnaire has been returned without creating any linkage to the questionnaire itself (Mangione 1995:69).

Web surveys create some unique ethical issues. It is easy for respondents to skip through a consent form when responding to a survey online and so never to consider the issues. A good practice is to provide a list of statements about the project and require that the respondent check a box to indicate acceptance of each one (Hewson et al. 2016:103). Screening for eligibility is another problem heightened in web surveys. A survey intended for adults could as easily be completed by children; in fact, more than 80% of children lie about their age when using social media sites. Recruiting respondents at websites that cater only to adults will reduce the problem, as will clear guidelines about eligibility criteria. Clear instructions can also minimize the problem of multiple submissions of a survey via the Internet (Toepoel 2016:48–49).

Confidentiality: A provision of research, in which identifying information that could be used to link respondents to their responses is available only to designated research personnel for specific research needs.

Anonymity: A provision of research, in which no identifying information is recorded that could be used to link respondents to their responses.

Conclusions

Survey research is an exceptionally efficient and productive method for investigating a wide array of social research questions. Mirowsky and Ross (2003) and Mirowsky (1999) were able to survey representative samples of Americans and older Americans and follow them for 6 years. These data allowed Mirowsky and Ross to investigate the relationships among education, social status, and health and how these relationships are changing.

In addition to the potential benefits for social science, considerations of time and expense frequently make a survey the preferred data collection method. One or more of the six survey designs reviewed in this chapter (including mixed mode) can be applied to almost any research question. It is no wonder that surveys have become the most popular research method in sociology and that they frequently inform discussion and planning about important social and political questions. As use of the Internet increases, survey research should become even more efficient and popular.

The relative ease of conducting at least some types of survey research leads many people to imagine that no particular training or systematic procedures are required. Nothing could be further from the truth. As a result of this widespread misconception, you will encounter a great many nearly worthless survey results. You must be prepared to examine carefully the procedures used in any survey before accepting its findings as credible. And if you decide to conduct a survey, you must be prepared to invest the time and effort that proper procedures require.

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Anonymity 304
Behavior coding 274
Bipolar response options 265
Cognitive interview 274
Computer-assisted personal interview (CAPI) 294
Computer-assisted telephone interview (CATI) 291
Confidentiality 304
Context effects 278
Contingent question 262
Cover letter 283
Double-barreled question 262
Double negative 261
Electronic survey 281
Fence-sitters 267
Filter question 262
Floaters 267
Forced-choice questions 268
Group-administered survey 286
Idiosyncratic variation 268
In-person interview 293
Interactive voice response (IVR) 292
Interpretive questions 275
Interview schedule 273
Labeled unipolar response options 265
Likert item 265
Mailed survey 283
Matrix questions 279
Mixed-mode survey 300
Omnibus survey 258
Part–whole question effects 278
Phone survey 286
Questionnaire 273
Skip pattern 262
Social desirability bias 265
Split-ballot design 258
Survey pretest 274
Survey research 255
Unlabeled unipolar response options 265
Web survey 296

Highlights

- Surveys are the most popular form of social research because of their versatility, efficiency, and generalizability. Many survey data sets, such as the GSS, are available for social scientists to use in teaching and research.
- Omnibus surveys cover a range of topics of interest and generate data useful to multiple sponsors.
- Survey designs must minimize the risk of errors of observation (measurement error) and errors of nonobservation (errors resulting from inadequate coverage, sampling error, and nonresponse). The likelihood of both types of error varies with the survey goals. For example, political polling can produce inconsistent results because of rapid changes in popular sentiment.
- Social exchange theory asserts that behavior is motivated by the return expected to the individual for the behavior. Survey designs must maximize the social rewards, minimize the costs of participating, and establish trust that the rewards will outweigh the costs.
- Questions must be worded carefully to avoid confusing respondents, encouraging a less-than-honest response, or triggering biases. Inclusion of “Don’t know” choices and neutral responses may help, but the presence of such options also affects the distribution of answers. Open-ended questions can be used to determine the meaning that respondents attach to their answers. Answers to any survey questions may be affected by the questions that precede them in a questionnaire or interview schedule.
- Sets of questions that comprise an index can reduce idiosyncratic variation in measurement of a concept. Indexes may be unidimensional or multidimensional. Responses to the questions in an index should be tested after data are collected to ensure that they can be combined as measures of a single concept, or of several related concepts, as intended.
- Questions can be tested and improved through review by experts, focus group discussions, cognitive interviews, behavior coding, and pilot testing. Every questionnaire and interview schedule should be pretested on a small sample that is like the sample to be surveyed.
- Interpretive questions should be used in questionnaires to help clarify the meaning of responses to critical questions.
- A survey questionnaire or interview schedule should be designed as an integrated whole, with each question and section serving some clear purpose and complementing the others.
- The cover letter for a mailed questionnaire should be credible, personalized, interesting, and responsible.
- Response rates in mailed surveys are typically well below 70% unless multiple mailings are made to nonrespondents and the questionnaire and cover letter are attractive, interesting, and carefully planned. Response rates for group-administered surveys are usually much higher.
- Phone interviews using random digit dialing allow fast turnaround and efficient sampling. Multiple callbacks are often required, and the rate of nonresponse to phone interviews is rising. Phone interviews should be no more than 30–45 minutes. Response rates to phone surveys have declined dramatically due to cell phones and caller ID.
- In-person interviews have several advantages over other types of surveys: They allow longer and more complex interview schedules, monitoring of the conditions when the questions are answered, probing for respondents’ understanding of the questions, and high response rates. However, the interviewer must balance the need to establish rapport with the respondent with the importance of maintaining control over the delivery of the interview questions.
- Electronic surveys may be e-mailed or posted on the web. Interactive voice response systems using the telephone are another option. At this time, use of the Internet is not sufficiently widespread to allow web surveys of the general population, but these approaches can be fast and efficient for populations with high rates of computer use.
- Mixed-mode surveys allow the strengths of one survey design to compensate for the weaknesses of another. However, questions and procedures must be designed carefully to reduce the possibility that responses to the same question will vary as a result of the mode of delivery.
- In deciding which survey design to use, researchers must consider the unique features and goals of the study. In general, in-person interviews are the strongest, but most expensive, survey design.
- Most survey research poses few ethical problems because respondents are able to decline to

participate—an option that should be stated clearly in the cover letter or introductory statement. Special care must be taken when questionnaires are administered in group settings (to “captive audiences”) and when sensitive personal questions are to be asked; subject confidentiality should always be preserved.

Discussion Questions

1. Response rates to phone surveys are declining, even as phone usage increases. Part of the problem is that lists of cell phone numbers are not available and wireless service providers may not allow outside access to their networks. Cell phone users may also have to pay for incoming calls. Do you think regulations should be passed to increase the ability of survey researchers to include cell phones in their random digit dialing surveys? How would you feel about receiving survey calls on your cell phone? What problems might result from "improving" phone survey capabilities in this way?
2. In-person interviews have for many years been the "gold standard" in survey research because the presence of an interviewer increases the response rate, allows better rapport with the interviewee, facilitates clarification of questions and instructions, and provides feedback about the interviewee's situation. However, researchers who design in-person interviewing projects are now making increasing use of technology to ensure consistent questioning of respondents and to provide greater privacy for respondents answering questions. But having a respondent answer questions on a laptop while the interviewer waits is a very different social process than asking the questions verbally. Which approach would you favor in survey research? What trade-offs might there be in quality of information collected, rapport building, and interviewee satisfaction?
3. Each of the following questions was used in a survey that I received in the past. Evaluate each question and its response choices using the guidelines for writing survey questions presented in this chapter. What errors do you find? Try to rewrite each question to avoid such errors and improve question wording.
    a. From an *InfoWorld* (computer publication) product evaluation survey:

        How interested are you in PostScript Level 2 printers?

        ____Very ____Somewhat ____Not at all

    b. From the Greenpeace National Marine Mammal Survey:

        Do you support Greenpeace's nonviolent, direct action to intercept whaling ships, tuna fleets, and other commercial fishermen to stop their wanton destruction of thousands of magnificent marine mammals?

        ____Yes ____No ____Undecided

    c. From a U.S. Department of Education survey of college faculty:

        How satisfied or dissatisfied are you with each of the following aspects of your instructional duties at this institution?

	Very	Somewhat	Somewhat	Very
	Dissat.	Dissat.	Satisf.	Satisf.
a. The authority I have to make decisions about what courses I teach	1	2	3	4
b. Time available for working with students as advisor, mentor	1	2	3	4

    d. From a survey about affordable housing in a Massachusetts community:

        Higher than single-family density is acceptable to make housing affordable.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
1	2	3	4	5

e. From a survey of faculty experience with ethical problems in research:

Are you reasonably familiar with the codes of ethics of any of the following professional associations?

	Very Familiar	Familiar	Not Too Familiar
American Sociological Association	1	2	0
Society for the Study of Social Problems	1	2	0
American Society of Criminology	1	2	0

If you are familiar with any of the above codes of ethics, to what extent do you agree with them?

Strongly Agree____ Agree____ No opinion____ Disagree____ Strongly Disagree____

Some researchers have avoided using a *professional code of ethics* as a guide for the following reasons. Which responses, if any, best describe your reasons for not using all or any of parts of the codes?

	Yes	No
1. Vagueness	1	0
2. Political pressures	1	0
3. Codes protect only individuals, not groups	1	0

f. From a survey of faculty perceptions:

Of the students you have observed while teaching college courses, please indicate the percentage who significantly improve their performance in the following areas.

Reading ____%

Organization ____%

Abstraction ____%

g. From a University of Massachusetts Boston student survey:

A person has a responsibility to stop a friend or relative from driving when drunk.

Strongly Agree____ Agree____ Disagree____ Strongly Disagree____

Even if I wanted to, I would probably not be able to stop most people from driving drunk.

Strongly Agree____ Agree____ Disagree____ Strongly Disagree____

Practice Exercises

1. Consider how you could design a split-ballot experiment to determine the effect of phrasing a question or its response choices in different ways. Check recent issues of the local newspaper for a question used in a survey of attitudes about a social policy or political position. Propose a hypothesis about how the wording of the question or its response choices might have influenced the answers people gave, and devise an alternative that differs only in this respect. Distribute these questionnaires to a large class (after your instructor makes the necessary arrangements) to test your hypothesis.
2. I received in my university mailbox some years ago a two-page questionnaire that began with the following cover letter at the top of the first page:

> Faculty Questionnaire
>
> This survey seeks information on faculty perception of the learning process and student performance in their undergraduate careers. Surveys have been distributed in universities in the Northeast, through random deposit in mailboxes of selected departments. This survey is being conducted by graduate students affiliated with the School of Education and the Sociology Department. We greatly appreciate your time and effort in helping us with our study.

Critique the “Faculty Questionnaire” cover letter, and then draft a more persuasive one.

3. Test your understanding of survey research terminology by completing one set of interactive exercises on survey design from the study site at edge.sagepub.com/schutt9e. Be sure to review the text on the pages indicated in relation to any answers you missed.
4. Review this chapter’s “Research That Matters” article on the book’s study site at edge.sagepub.com/schutt9e. Describe the sampling and measurement methods used and identify both strong and weak points of the survey design. Would a different type of survey design (in-person, phone, mailed, web) have had any advantages? Explain your answer.

Ethics Questions

1. Group-administered surveys are easier to conduct than other types of surveys, but they always raise an ethical dilemma. If a teacher allows a social research survey to be distributed in his or her class, or if an employer allows employees to complete a survey on company time, is the survey truly voluntary? Is it sufficient to read a statement to the group members stating that their participation is entirely up to them? How would you react to a survey in your class? What general guidelines should be followed in such situations?
2. Patricia Tjaden and Nancy Thoennes (2000) sampled adults with random digit dialing to study violent victimization from a nationally representative sample of adults. What ethical dilemmas do you see in reporting victimizations that are identified in a survey? What about when the survey respondents are under the age of 18? What about children under the age of 12?

Web Exercises

1. Who does survey research and how do they do it? These questions can be answered through careful inspection of ongoing surveys and the organizations that administer them at www.ciser.cornell.edu/info/polls.shtml. Spend some time reading about the different survey research organizations, and write a brief summary of the types of research they conduct, the projects in which they are involved, and the resources they offer on their websites. What are the distinctive features of different survey research organizations?
2. Go to the Research Triangle Institute site at www.rti.org. Click on “Services and Capabilities” and then “Surveys and Data Collection.” Read some of the stories about survey design, instrument development, survey methodologies, and data collection. Which innovations in survey design can you identify?
3. Go to the UK Data Service at http://discover.ukdataservice.ac.uk/variables. In the search box, enter topics of interest such as “health” or “inequality.” Review five questions for two topic areas and critique them in terms of the principles for question writing that you have learned. Do you find any question features that might be attributed to the use of British English?

Video Interview Questions

Listen to the researcher interview for Chapter 8 at edge.sagepub.com/schutt9e.

1. What two issues should survey researchers consider when designing questions?
2. Why is cognitive testing of questions important?

SPSS Exercises

What can we learn from the General Social Survey (GSS2016) data about the orientations of people who support capital punishment? Is it related to religion? Reflective of attitudes toward race? What about political views? Is it a guy thing? Do attitudes and behavior concerning guns have some relation to support for capital punishment?

1. To answer these questions, we will use some version of each of the following variables in our analysis: PARTYID3, GUNLAW, HELPBLK, DISCAFFF, FUND, OWNGUN, and CAPPUN. Check the wording of each of these questions at the University of Michigan's GSS website (click on "Browse GSS Variables" and use the mnemonic listing of variables to find those in the list above): www.norc.org/GSS+Website

    How well does each of these questions meet the guidelines for writing survey questions? What improvements would you suggest?
2. Now generate cross-tabulations to show the relationship between each of these variables, treated as independent variables, and support for capital punishment. A cross-tabulation can be used to display the distribution of responses on the dependent variable for each category of the independent variable. For this purpose, you should substitute several slightly different versions of the variables you just reviewed. From the menu, select Analyze/Descriptive Statistics/Crosstabs:

    Rows: CAPPUN
    Columns: SEX, PARTYID3, GUNLAW, HELPBLK, DISCAFF, FUND, OWNGUN
    Cells: column percentages

    (If you have had a statistics course, you will also want to request the chi-square statistic for each of these tables.)

    Describe the relationship you have found in the tables, noting the difference in the distribution of the dependent (row) variable—support for capital punishment—between the categories of each of the independent (column) variables.
3. Summarize your findings. What attitudes and characteristics are associated strongly with support for the death penalty?
4. What other hypotheses would you like to test? What else do you think needs to be considered to help you understand the relationships you have identified? For example, should you consider the race of the respondents? Why or why not?
5. Let's take a minute to learn about recoding variables. If you generate the frequencies for POLVIEWS and for POLVIEWS3, you'll see how I recoded POLVIEWS3. Why? Because I wanted to use a simple categorization by political party views in the cross-tabulation. You can try to replicate my recoding in SPSS. From the menu, click Transform/Recode/Into different variables. Identify the old variable name and type in the new one. Type in the appropriate sets of old values and the corresponding new values. You may need to check the numerical codes corresponding to the old values with the variable list pull-down menu (the ladder icon with a question mark).

Developing a Research Proposal

These steps focus again on the "Research Design" decisions, but this time assuming that you will use a survey design (Exhibit 3.10, #13 to #17).

1. Write 10 questions for a one-page questionnaire that concerns your proposed research question. Your questions should operationalize at least three of the variables on which you have focused, including at least one independent and one dependent variable (you may have multiple questions to measure some variables). Make all but one of your questions closed-ended. If you completed the "Developing a Research Proposal" exercises in Chapter 4, you can select your questions from the

ones you developed for those exercises.

2. Conduct a preliminary pretest of the questionnaire by conducting cognitive interviews with two students or other persons like those to whom the survey is directed. Follow up the closed-ended questions with open-ended probes that ask the students what they meant by each response or what came to mind when they were asked each question. Account for the feedback you receive when you revise your questions.
3. Polish the organization and layout of the questionnaire, following the guidelines in this chapter. Prepare a rationale for the order of questions in your questionnaire. Write a cover letter directed to the appropriate population that contains appropriate statements about research ethics (human subjects' issues).

Chapter 9 Quantitative Data Analysis

- Research That Matters, Questions That Count
- Introducing Statistics
  - Case Study: The Likelihood of Voting
- Preparing for Data Analysis
- Displaying Univariate Distributions
  - Graphs
  - Frequency Distributions
    - *Ungrouped Data*
    - *Grouped Data*
    - *Combined and Compressed Distributions*
- Summarizing Univariate Distributions
- Research in the News: Why Key State Polls Were Wrong About Trump
  - Measures of Central Tendency
    - *Mode*
    - *Median*
    - *Mean*
    - *Median or Mean?*
  - Measures of Variation
    - *Range*
    - *Interquartile Range*
    - *Variance*
    - *Standard Deviation*
- Analyzing Data Ethically: How Not to Lie With Statistics
- Cross-Tabulating Variables
  - Constructing Contingency Tables
  - Graphing Association
  - Describing Association
  - Evaluating Association
  - Controlling for a Third Variable
    - *Intervening Variables*
    - *Extraneous Variables*
    - *Specification*
- Careers and Research
- Regression Analysis
- Performing Meta-Analyses
  - Case Study: Patient–Provider Race Concordance and Minority Health Outcomes
- Analyzing Data Ethically: How Not to Lie About Relationships

Conclusions

Research That Matters, Questions That Count

Does college influence sociopolitical attitudes? It's one of those questions that has generated passionate political debate, as a growing body of sociological research demonstrates that simplistic answers are likely to be wrong. The basic problem is this: College graduates in the United States are more liberal than others, but this could be due to liberals being more likely to go to college rather than to students' experiences while in college. Colin Campbell at the University of Wisconsin–Madison and Jonathan Horowitz at the University of North Carolina at Chapel Hill decided to investigate this question by analyzing data collected with the General Social Survey (GSS) and the Study of American Families (SAF).

Previous researchers who have tried to distinguish selection effects from college effects have taken account of family socioeconomic status, but they have not often been able to take account of parents' sociopolitical attitudes. Campbell and Horowitz dealt with the problem by studying respondents in the 1994 GSS who had indicated they had at least one sibling, and combining their survey responses with comparable data from their siblings that was collected in the SAF. The analysis identifies some differences between those who attended college and their siblings who did not. Earning a college degree increased support for civil liberties and egalitarian gender roles, but the greater political liberalism of college graduates seems largely due to their family background.

1. What causal mechanism do you think might account for an effect of college on sociopolitical attitudes? Compare your reasoning about mechanisms to that of Campbell and Horowitz (pp. 41–42).
2. Do you think that focusing their comparison on siblings solves the problem of distinguishing selection effects from college effects? What problems do you think might occur with this approach? Compare your concerns to those discussed by Campbell and Horowitz (pp. 44, 55–56).

In this chapter, you will learn the basic statistical tools used to describe variation in variables and the relations between them, as well as some of the findings about influences on voting and political attitudes. By the end of the chapter, you will understand the primary steps involved in the analysis of quantitative data and some of the potential pitfalls in such analyses. As you read the chapter, extend your understanding by reading the 2016 *Sociology of Education* article by Colin Campbell and Jonathan Horowitz at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 9 at edge.sagepub.com/schutt9e.

Source: Campbell, Colin and Jonathan Horowitz. 2016. "Does College Influence Sociopolitical Attitudes?" *Sociology of Education* 89(1):40–58.

This chapter introduces several common statistics used in social research and highlights the factors that must be considered when using and interpreting statistics. Think of it as a review of fundamental social statistics, if you have already studied them, or as an introductory overview, if you have not. Two preliminary sections lay the foundation for studying statistics. In the first, I discuss the role of statistics in the research process, returning to themes and techniques with which you are already familiar. In the second preliminary section, I outline the process of preparing data for statistical analysis. In the rest of the chapter, I explain how to describe the distribution of single variables and the relationship between variables. Along the way, I address ethical issues related to data analysis. This chapter will have been successful if it encourages you to use statistics responsibly, to evaluate statistics critically, and to seek opportunities for extending your statistical knowledge.

Although many colleges and universities offer social statistics in a separate course, and for good reason (there's a *lot* to learn), I don't want you to think of this chapter as something that deals with a different topic than the rest of this book. Data analysis is an integral component of research methods, and it's important that any proposal for quantitative research include a plan for the data analysis that will follow data collection. You have to anticipate your data analysis needs if you expect your research design to secure the requisite data.

Introducing Statistics

Statistics play a key role in achieving valid research results—in measurement, causal validity, and generalizability. Some statistics are useful primarily to describe the results of measuring single variables and to construct and evaluate multi-item scales. These statistics include frequency distributions, graphs, measures of central tendency and variation, and reliability tests. Other statistics are useful primarily in achieving causal validity, by helping us describe the association between variables and control for, or otherwise account for, other variables. This chapter introduces cross-tabulation as a technique for measuring association and controlling other variables. All such statistics are termed descriptive statistics because they are used to describe the distribution of, and relationship between, variables.

You have already learned in Chapter 5 that it is possible to estimate the degree of confidence that can be placed in generalization from a sample to the population from which the sample was selected. The statistics used in making these estimates are termed *inferential statistics.* In this chapter, I introduce the use of inferential statistics for testing hypotheses involving sample data.

Social theory and the results of prior research should guide our statistical choices, as they guide the choice of other research methods. There are so many particular statistics and so many ways for them to be used in data analysis that even the best statistician can be lost in a sea of numbers if he or she does not use prior research and theorizing to develop a coherent analysis plan. It is also important to choose statistics that are appropriate to the level of measurement of the variables to be analyzed. As you learned in Chapter 4, numbers used to represent the values of variables may not signify different quantities, meaning that many statistical techniques will be inapplicable for some variables.

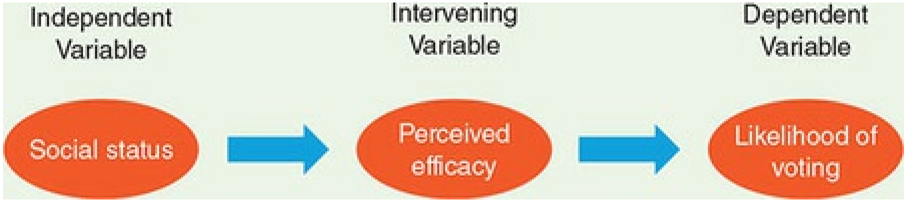
Descriptive statistics: Statistics used to describe the distribution of and relationship between variables.

Case Study: The Likelihood of Voting

In this chapter, I use for examples data from the 2016 General Social Survey (GSS) on voting and the variables associated with it (Smith et al. 2017), and I will focus on a research question about political participation: What influences the likelihood of voting? Prior research on voting in both national and local settings provides a great deal of support for one hypothesis: The likelihood of voting increases with social status (Manza, Brooks, and Sauder 2005:208; Milbrath and Goel 1977:92–95; Salisbury 1975:326; Verba and Nie 1972:892). Research suggests that social status influences the likelihood of voting through the intervening variable of perceived political efficacy, or the feeling that one's vote matters (see Exhibit 9.1). But some research findings on political participation are inconsistent with the social status–voting hypothesis. For example, African Americans participate in politics at higher rates than do white Americans of similar social status—at least when there is an African American candidate for whom to vote (Manza et al. 2005:209; Verba and Nie 1972; Verba, Nie, and Kim 1978). This discrepant finding suggests that the impact of social status on voting and other forms of political participation varies with the social characteristics of potential participants.

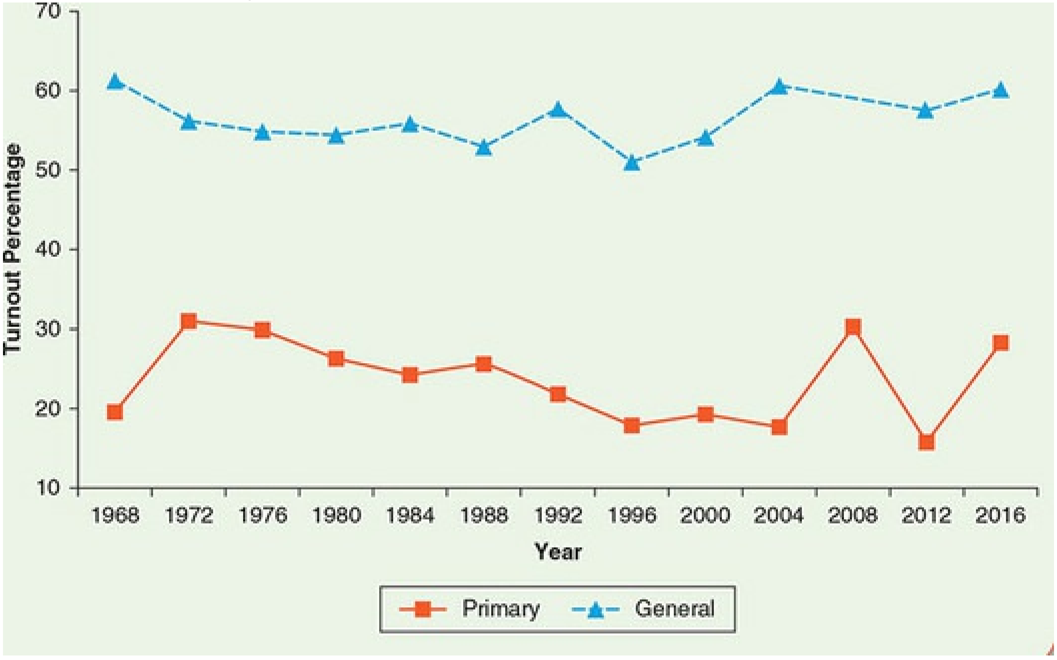
The rate of voting of 60.2% in the 2016 presidential election came close to matching the recent high of 62.3% in the election of Barack Obama in 2008 (see Exhibit 9.2) (Gans 2008; Liptak 2012; Wilson 2017). Participation among African Americans continued to increase and for the first time in 2012 surpassed the participation rate for white non-Hispanic Americans, although turnout remained much lower among Hispanics and Asian Americans. Turnout dropped among young people in the 2012 and 2016 presidential elections compared with the historic high reached in the 2008 election (CIRCLE 2013; Regan 2016).

Exhibit 9.1 Causal Model of Likelihood of Voting



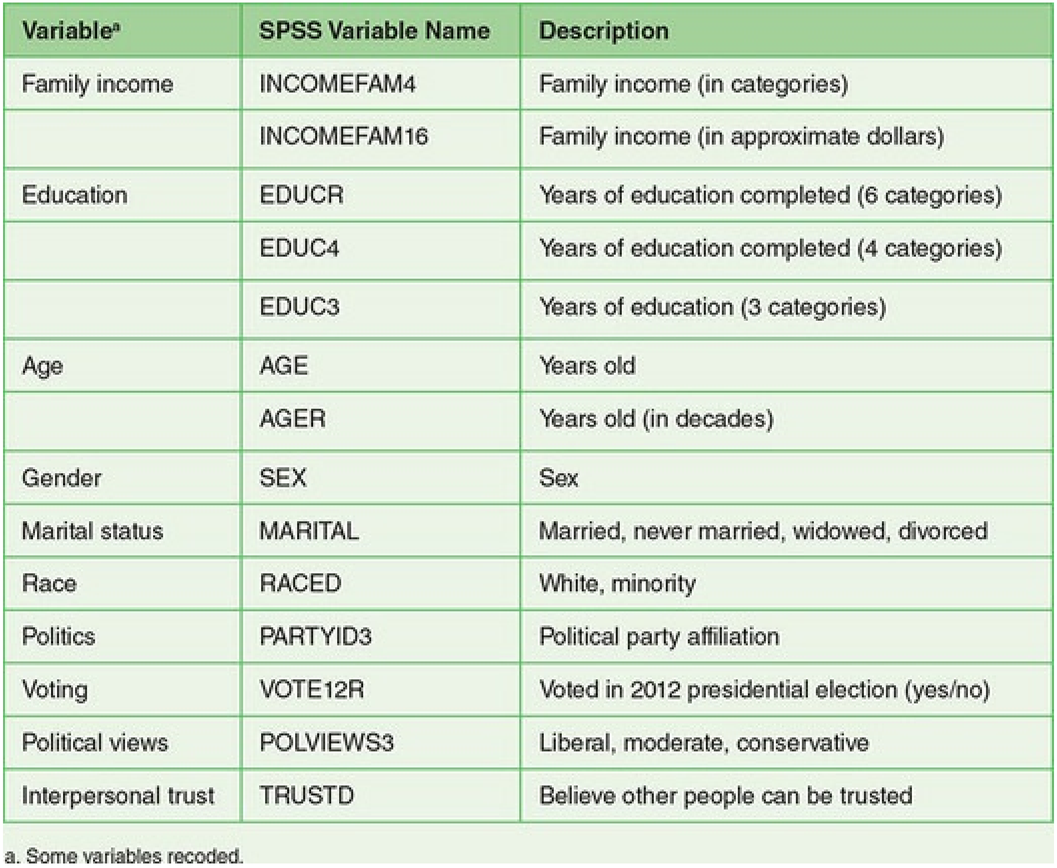
If we are guided by prior research, a test of the hypothesis that likelihood of voting increases with social status should also account for political efficacy and some social characteristics, such as race. We can find indicators for each of these variables, except political efficacy, in the 2016 GSS (see Exhibit 9.3). We will substitute the variable *interpersonal trust* for political efficacy. I will use these variables to illustrate particular statistics throughout this chapter, drawing on complete 2016 GSS data. You can replicate my analysis with the 2016x GSS data set posted on the study site for this book at edge.sagepub.com/schutt9e.

Exhibit 9.2 Voting in Presidential Primaries and General Elections, 1968–2016



Source: Center for the Study of the American Electorate, American University, Preliminary Primary Turnout Report.

Exhibit 9.3 List of GSS 2016 Variables for Analysis of Voting



Variable[a]	SPSS Variable Name	Description
Family income	INCOMEFAM4	Family income (in categories)
	INCOMEFAM16	Family income (in approximate dollars)
Education	EDUCR	Years of education completed (6 categories)
	EDUC4	Years of education completed (4 categories)
	EDUC3	Years of education (3 categories)
Age	AGE	Years old
	AGER	Years old (in decades)
Gender	SEX	Sex
Marital status	MARITAL	Married, never married, widowed, divorced
Race	RACED	White, minority
Politics	PARTYID3	Political party affiliation
Voting	VOTE12R	Voted in 2012 presidential election (yes/no)
Political views	POLVIEWS3	Liberal, moderate, conservative
Interpersonal trust	TRUSTD	Believe other people can be trusted

a. Some variables recoded.

Source: General Social Survey, National Opinion Research Center 2016.

Preparing for Data Analysis

My analysis of voting in this chapter is an example of secondary data analysis, which you will learn about in Chapter 14. Using secondary data in this way has a major advantage: The researcher doesn't have to secure the funds and spend the time required to collect his or own data. But there are also disadvantages: If you did not design the study yourself, it is unlikely that all the variables that you think should have been included actually were included and were measured in the way that you prefer. In addition, the sample may not represent just the population in which you are interested, and the study design may be only partially appropriate to your research question. For example, because it is a survey of individuals, the GSS lacks measures of political context (such as the dominant party in an area). Because the survey sample is selected only from the United States and because the questions concern just one presidential election, we will not be able to address directly the larger issues of political context that are represented in cross-national and longitudinal research (for more on cross-national and longitudinal research, see Verba et al. 1978).

If you conduct your own survey or experiment, your quantitative data must be prepared in a format suitable for computer entry. Several options are available. Questionnaires or other data entry forms can be designed for scanning or direct computer entry (see Exhibit 9.4). Coding of all responses should be done before data entry by assigning each a unique numerical value. Once the computer database software is programmed to recognize the response codes, the forms can be fed through a scanner and the data will then be entered directly into the database. If responses or other forms of data have been entered on nonscannable paper forms, a computer data entry program should be used that will allow the data to be entered into the databases by clicking on boxes corresponding to the response codes. Alternatively, if a data entry program is not used, responses can be typed directly into a computer database. If data entry is to be done this way, the questionnaires or other forms should be precoded. Precoding means that a number represents every response choice, and respondents are instructed to indicate their response to a question by checking a number. It will then be easier to type in the strings of numbers than to type in the responses themselves.

Data entry: The process of typing (word processing) or otherwise transferring data on survey or other instruments into a computer file.

Coding: The process of assigning a unique numerical code to each response to survey questions.

Precoding: The process through which a questionnaire or other survey form is prepared so that a number represents every response choice, and respondents are instructed to indicate their response to a question by checking a number.

Whatever data entry method is used, the data must be checked carefully for errors—a process called data cleaning. The first step in data cleaning is to check responses before they are entered into the database to make sure that one and only one valid answer code has been clearly circled or checked for each question (unless multiple responses are allowed or a skip pattern was specified). Written answers can be assigned their own numerical codes. The next step in data cleaning is to make sure that no invalid codes have been entered. Invalid codes are codes that fall outside the range of allowable values for a given variable and those that represent impossible combinations of responses to two or more questions. (For example, if a respondent says that he or she did not vote in an election, a response to a subsequent question indicating whom that person voted for would be invalid.) Most survey research organizations now use a database management program to control data entry. The program prompts the data entry clerk for each response code, checks the code to ensure that it represents a valid response for that variable, and saves the response code in the data file. This process reduces sharply the possibility of data entry errors.

If data are typed into a text file or entered directly through the data sheet of a statistics program, a computer program must be written to "define the data." A data definition program identifies the variables that are coded in each column or range of columns, attaches meaningful labels to the codes, and distinguishes values representing missing data. The procedures for doing so vary with the specific statistical package used. I used the Statistical Package for the Social Sciences (SPSS) for the analysis in this chapter; you will find examples of SPSS commands for defining and analyzing data in Appendix D (on the study site at edge.sagepub.com/schutt9e). More information on using SPSS is contained in SPSS manuals and in the SAGE Publications volume *Using IBM SPSS Statistics for Social Statistics and Research Methods,* 3rd edition, by William E. Wagner III (2011).

Data cleaning: The process of checking data for errors after the data have been entered in a computer file.

Displaying Univariate Distributions

The first step in data analysis is usually to display the variation in each variable of interest. For many descriptive purposes, the analysis may go no further. Graphs and frequency distributions are the two most popular approaches; both allow the analyst to display the distribution of cases across the categories of a variable. Graphs have the advantage of providing a picture that is easier to comprehend, although frequency distributions are preferable when exact numbers of cases having particular values must be reported and when many distributions must be displayed in a compact form.

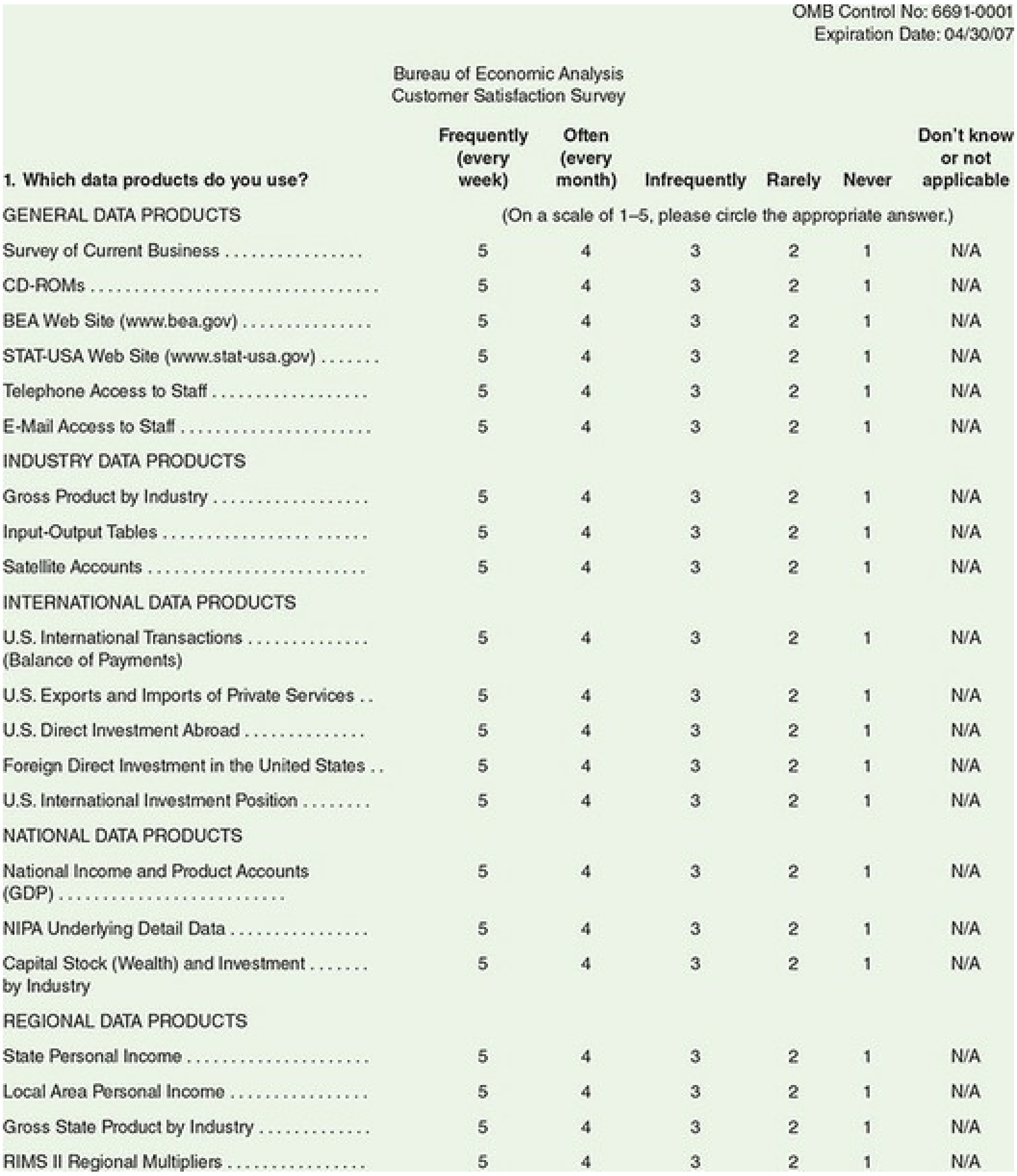
Whichever type of display is used, the primary concern of the data analyst is to display accurately the distribution's shape, that is, to show how cases are distributed across the values of the variable. Three features of shape are important: central tendency, variability, and skewness (lack of symmetry). All three features can be represented in a graph or in a frequency distribution.

Central tendency: The most common value (for variables measured at the nominal level) or the value around which cases tend to center (for a quantitative variable).

Variability: The extent to which cases are spread out through the distribution or clustered in just one location.

Skewness: The extent to which cases are clustered more at one or the other end of the distribution of a quantitative variable rather than in a symmetric pattern around its center. Skew can be positive (a right skew), with the number of cases tapering off in the positive direction, or negative (a left skew), with the number of cases tapering off in the negative direction.

Exhibit 9.4 Form for Direct Data Entry



OMB Control No: 6691-0001
Expiration Date: 04/30/07

Bureau of Economic Analysis
Customer Satisfaction Survey

1. Which data products do you use?	**Frequently (every week)**	**Often (every month)**	**Infrequently**	**Rarely**	**Never**	**Don't know or not applicable**
GENERAL DATA PRODUCTS	(On a scale of 1–5, please circle the appropriate answer.)					
Survey of Current Business	5	4	3	2	1	N/A
CD-ROMs	5	4	3	2	1	N/A
BEA Web Site (www.bea.gov)	5	4	3	2	1	N/A
STAT-USA Web Site (www.stat-usa.gov)	5	4	3	2	1	N/A
Telephone Access to Staff	5	4	3	2	1	N/A
E-Mail Access to Staff	5	4	3	2	1	N/A
INDUSTRY DATA PRODUCTS						
Gross Product by Industry	5	4	3	2	1	N/A
Input-Output Tables	5	4	3	2	1	N/A
Satellite Accounts	5	4	3	2	1	N/A
INTERNATIONAL DATA PRODUCTS						
U.S. International Transactions (Balance of Payments)	5	4	3	2	1	N/A
U.S. Exports and Imports of Private Services	5	4	3	2	1	N/A
U.S. Direct Investment Abroad	5	4	3	2	1	N/A
Foreign Direct Investment in the United States	5	4	3	2	1	N/A
U.S. International Investment Position	5	4	3	2	1	N/A
NATIONAL DATA PRODUCTS						
National Income and Product Accounts (GDP)	5	4	3	2	1	N/A
NIPA Underlying Detail Data	5	4	3	2	1	N/A
Capital Stock (Wealth) and Investment by Industry	5	4	3	2	1	N/A
REGIONAL DATA PRODUCTS						
State Personal Income	5	4	3	2	1	N/A
Local Area Personal Income	5	4	3	2	1	N/A
Gross State Product by Industry	5	4	3	2	1	N/A
RIMS II Regional Multipliers	5	4	3	2	1	N/A

Source: U.S. Bureau of Economic Analysis (2004):14.

These features of a distribution's shape can be interpreted in several different ways, and they are not all appropriate for describing every variable. In fact, all three features of a distribution can be distorted if graphs, frequency distributions, or summary statistics are used inappropriately.

A variable's level of measurement is the most important determinant of the appropriateness of particular statistics. For example, we cannot talk about the skewness (lack of symmetry) of a variable measured at the nominal level. If the values of a variable cannot be ordered from lowest or highest—if the ordering of the values is arbitrary—we cannot say that the distribution is not symmetric because we could just reorder the values to make the distribution more (or less) symmetric. Some measures of central tendency and variability are also inappropriate for variables measured at the nominal level.

The distinction between variables measured at the ordinal level and those measured at the interval or ratio level should also be considered when selecting statistics for use, but social researchers differ in just how much importance they attach to this distinction. Many social researchers think of ordinal variables as imperfectly measured interval-level variables and believe that, in most circumstances, statistics developed for interval-level variables also provide useful summaries for ordinal variables. Other social researchers believe that variation in ordinal variables will often be distorted by statistics that assume an interval level of measurement. We will touch on some of the details in the following sections on particular statistical techniques.

We will now examine graphs and frequency distributions that illustrate these three features of shape. Summary statistics used to measure specific aspects of central tendency and variability are presented in a separate section. There is a summary statistic for the measurement of skewness, but it is used only rarely in published research reports and will not be presented here.

Graphs

A picture often is worth some unmeasurable quantity of words. Even for the uninitiated, graphs can be easy to read, and they highlight a distribution's shape. They are useful particularly for exploring data because they show the full range of variation and identify data anomalies that might be in need of further study. And good, professional-looking graphs can now be produced relatively easily with software available for personal computers. There are many types of graphs, but the most common and most useful are bar charts, histograms, and frequency polygons. Each has two axes, the vertical axis (the *y*-axis) and the horizontal axis (the *x*-axis), and labels to identify the variables and the values, with tick marks showing where each indicated value falls along the axis.

A bar chart contains solid bars separated by spaces. It is a good tool for displaying the distribution of variables measured at the nominal level because there is, in effect, a gap between each of the categories. The bar chart of marital status in Exhibit 9.5 indicates that almost half of adult Americans were married at the time of the survey. Smaller percentages were divorced, separated, or widowed, and more than one quarter had never married. The most common value in the distribution is married, so this would be the distribution's central tendency. There is a moderate amount of variability in the distribution because the half that are not married are spread across the categories of widowed, divorced, separated, and never married. Because marital status is not a quantitative variable, the order in which the categories are presented is arbitrary, and so skewness is not relevant.

Histograms, in which the bars are adjacent, are used to display the distribution of quantitative variables that vary along a continuum that has no necessary gaps. Exhibit 9.6 shows a histogram of years of education from the 2016 GSS data. The distribution has a clump of cases at 12 years—about one third of the total. The distribution is negatively skewed because the percentage of cases tapers off to the low end much more quickly, with a long tail, compared with the shape of the distribution toward the high end.

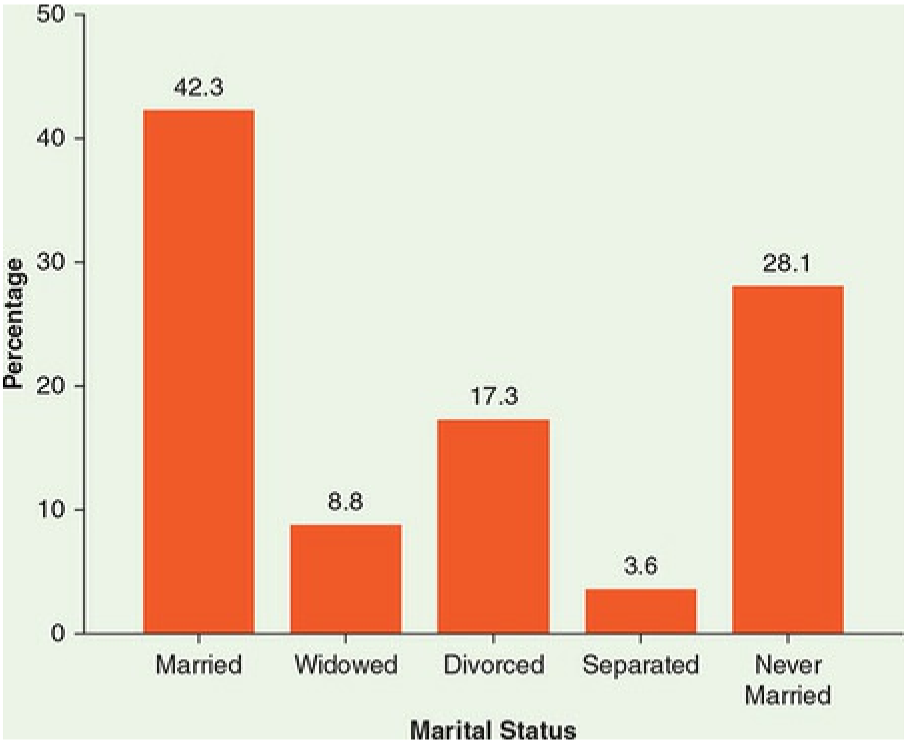
Bar chart: A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces.

Histogram: A graphic for quantitative variables in which the variable's distribution is displayed with adjacent bars.

In a frequency polygon, a continuous line connects the points representing the number or percentage of cases with each value. The frequency polygon is an alternative to the histogram when the distribution of a quantitative variable must be displayed; this alternative is particularly useful when the variable has a wide range of values. It is easy to see in the frequency polygon of years of education in Exhibit 9.7 that the most common value is 12 years, high school completion, and that this value also seems to be at the center of the distribution. There is moderate variability in the distribution, with many cases having more than 12 years of education and almost one third having completed at least 4 years of college (16 years). The distribution is highly skewed in the negative direction, with few respondents reporting fewer than 10 years of education.

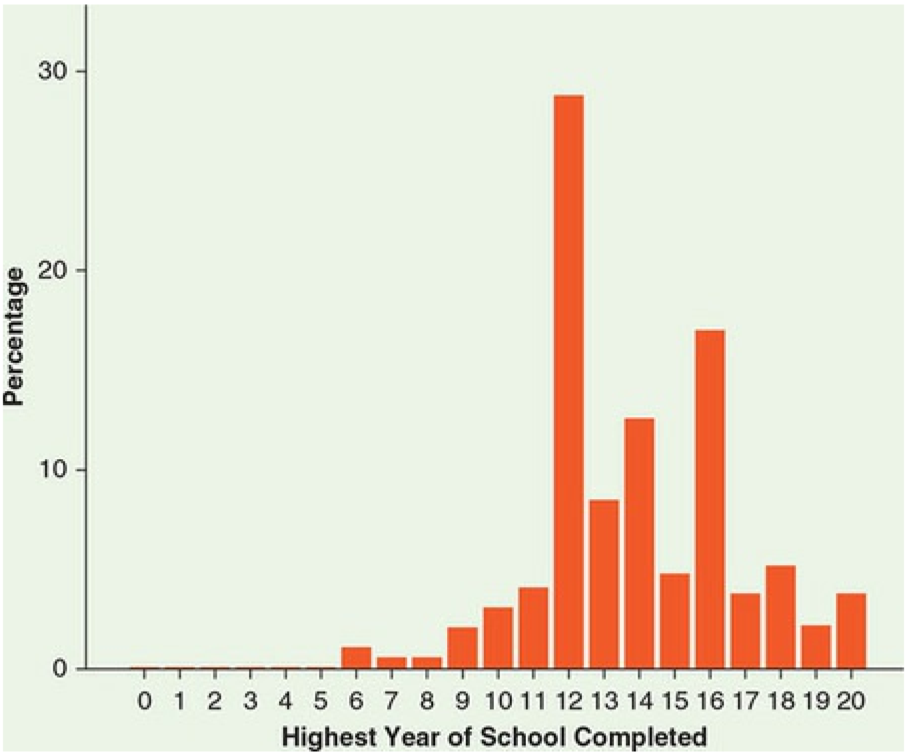
Frequency polygon: A graphic for quantitative variables in which a continuous line connects data points representing the variable's distribution.

Exhibit 9.5 Bar Chart of Marital Status



Source: General Social Survey National Opinion Research Center 2016.

Exhibit 9.6 Histogram of Years of Education



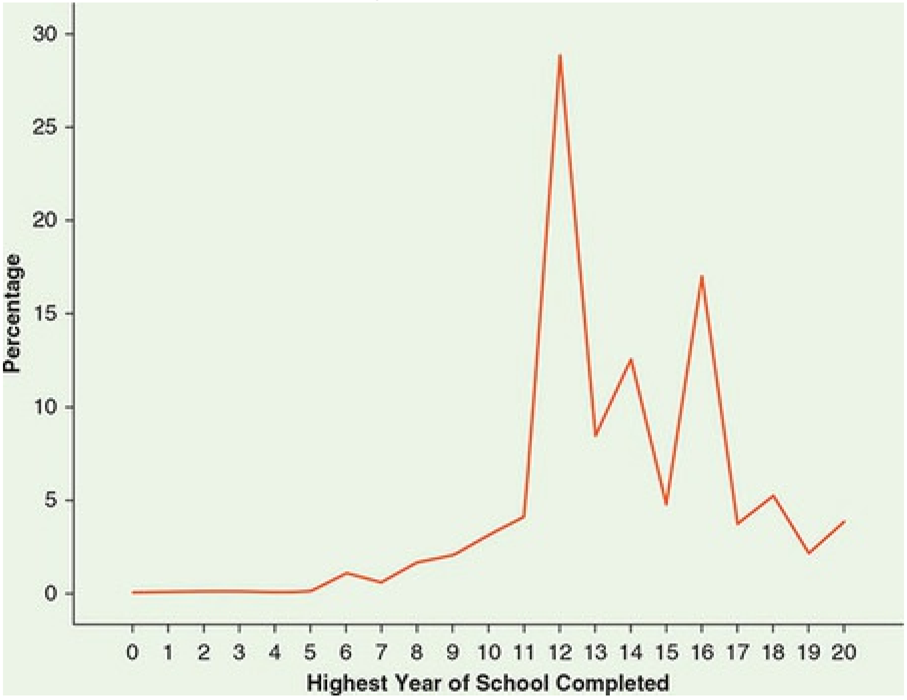
Source: General Social Survey National Opinion Research Center 2016.

If graphs are misused, they can distort, rather than display, the shape of a distribution. Compare, for example, the two graphs in Exhibit 9.8. The first graph shows that high school seniors reported relatively stable rates of lifetime use of cocaine between 1980 and 1985. The second graph, using exactly the same numbers, appeared in a 1986 *Newsweek* article on the coke plague (Orcutt and Turner 1993). Looking at this graph, you would think that the rate of cocaine usage among high school seniors had increased dramatically during this period. But, in fact, the difference between the two graphs is due simply to changes in how the graphs are drawn. In the "plague graph" the percentage scale on the vertical axis begins at 15 rather than at 0, making what was about a 1-percentage point increase look very big indeed. In addition, omission from the plague graph of the more rapid increase in reported usage between 1975 and 1980 makes it look as if the tiny increase in 1985 were a new, and thus more newsworthy, crisis.

Adherence to several guidelines (Tufte 1983; Wallgren et al. 1996) will help you spot these problems and avoid them in your own work:

- The difference between bars can be exaggerated by cutting off the bottom of the vertical axis and displaying less than the full height of the bars. Instead, begin the graph of a quantitative variable at 0 on both axes. It may be reasonable, at times, to violate this guideline, as when an age distribution is presented for a sample of adults, but in this case be sure to mark the break clearly on the axis.

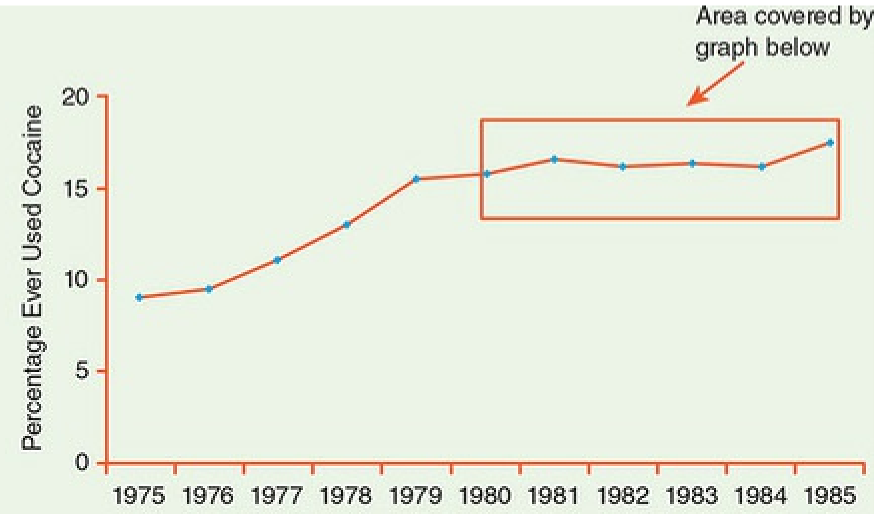
Exhibit 9.7 Frequency Polygon of Years of Education



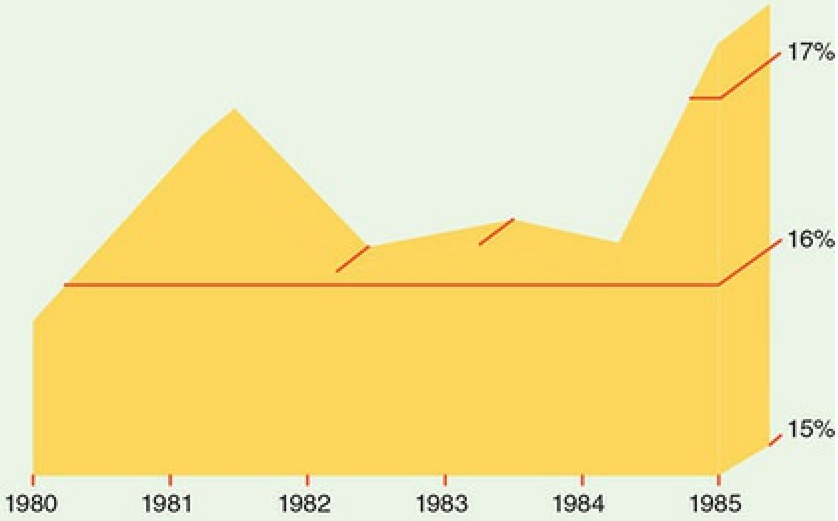
Source: General Social Survey National Opinion Research Center 2016.

- Bars of unequal width, including pictures instead of bars, can make particular values look as if they carry more weight than their frequency warrants. Always use bars of equal width.
- Either shortening or lengthening the vertical axis will obscure or accentuate the differences in the number of cases between values. The two axes usually should be of approximately equal length.
- Avoid chart junk that can confuse the reader and obscure the distribution's shape (a lot of verbiage or umpteen marks, lines, lots of cross-hatching, etc.).

Exhibit 9.8 Two Graphs of Cocaine Usage



A. University of Michigan Institute for Social Research
Time Series for Lifetime Prevalence of Cocaine Use



B. Final Stages of Construction

Source: Adapted from Orcutt and Turner (1993). Copyright 1993 by the Society for the Study of Social Problems. Reprinted by permission.

Frequency Distributions

A frequency distribution displays the number of cases, percentage (the relative frequencies) of cases, or both, corresponding to each of a variable's values or group of values. The components of the frequency distribution should be clearly labeled, with a title, a stub (labels for the values of the variable), a caption (identifying whether the distribution includes frequencies, percentages, or both), and perhaps the number of missing cases. If percentages, rather than frequencies, are presented (sometimes both are included), the total number of cases in the distribution (the base number *N*) should be indicated (see Exhibit 9.9).

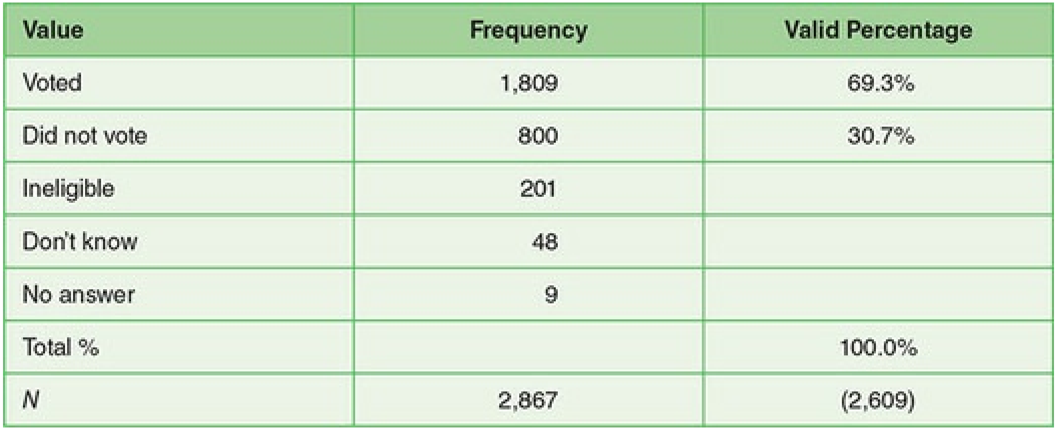
Frequency distribution: A numerical display showing the number of cases, and usually the percentage of cases (the relative frequencies), corresponding to each value or group of values of a variable

Base number (N): The total number of cases in a distribution.

Ungrouped Data

Constructing and reading frequency distributions for variables with few values is not difficult. The frequency distribution of voting in Exhibit 9.9, for example, shows that 69.3% of the respondents eligible to vote said they voted, and 30.7% reported they did not vote. The total number of respondents to this question was 2,609, although 2,867 actually were interviewed. The rest were ineligible to vote, said they did not know whether they had voted or not, or gave no answer.

Exhibit 9.9 Frequency Distribution of Voting in the 2012 Election



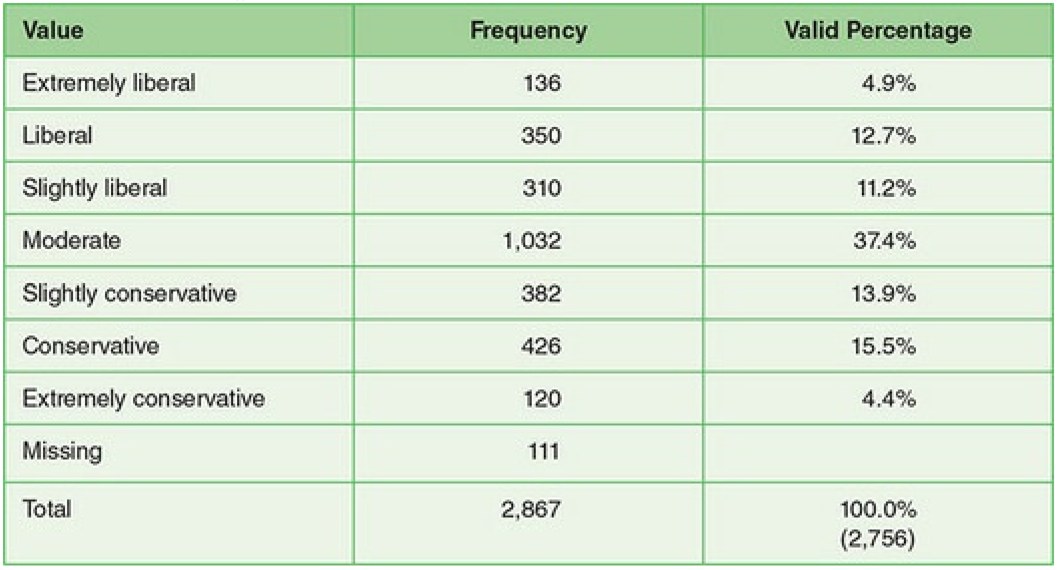
Value	**Frequency**	**Valid Percentage**
Voted	1,809	69.3%
Did not vote	800	30.7%
Ineligible	201	
Don't know	48	
No answer	9	
Total %		100.0%
N	2,867	(2,609)

Source: General Social Survey (National Opinion Research Center 2016).

Political ideology was measured with a question having seven response choices, resulting in a longer but still relatively simple frequency distribution (see Exhibit 9.10). The most common response was "moderate," with 37.4% of the sample that responded choosing this label to represent their political ideology. The distribution has a symmetric shape, although with somewhat more respondents identifying themselves as conservative rather than liberal.

If you compare Exhibits 9.10 and 9.6, you can see that a frequency distribution (Exhibit 9.10) can provide more precise information than a graph (Exhibit 9.6) about the number and percentage of cases in a variable's categories. Often, however, it is easier to see the shape of a distribution when it is graphed. When the goal of a presentation is to convey a general sense of a variable's distribution, particularly when the presentation is to an audience not trained in statistics, the advantages of a graph outweigh those of a frequency distribution.

Exhibit 9.10 Frequency Distribution of Political Views



Value	Frequency	Valid Percentage
Extremely liberal	136	4.9%
Liberal	350	12.7%
Slightly liberal	310	11.2%
Moderate	1,032	37.4%
Slightly conservative	382	13.9%
Conservative	426	15.5%
Extremely conservative	120	4.4%
Missing	111	
Total	2,867	100.0% (2,756)

Source: General Social Survey National Opinion Research Center 2016.

Grouped Data

Many frequency distributions (and graphs) require grouping of some values after the data are collected. There are two reasons for grouping:

1. There are more than 15–20 values to begin with, a number too large to be displayed in an easily readable table.
2. The distribution of the variable will be clearer or more meaningful if some of the values are combined.

Inspection of Exhibit 9.11 should clarify these reasons. In the first distribution, which is only a portion of the entire ungrouped GSS age distribution, it is very difficult to discern any shape, much less the central tendency. In the second distribution, age is grouped in the familiar 10-year intervals (except for the first, abbreviated category), and the distribution's shape is immediately clear.

Once we decide to group values, or categories, we have to be sure that in doing so, we do not distort the distribution. Adhering to the following guidelines for combining values in a frequency distribution will prevent many problems:

- Categories should be logically defensible and preserve the distribution's shape.
- Categories should be mutually exclusive and exhaustive, so that every case should be classifiable in one and only one category.

Violating these two guidelines is easier than you might think. If you were to group all the ages above 59 together, as 60 or higher, it would create the appearance of a bulge at the high end of the age distribution, with 30.1% of the cases. Combining other categories so that they include a wide range of values could create the same type of misleading impression. In some cases, however, the most logically defensible categories will vary in size. A good example would be grouping years of education as less than 8 (did not finish grade school), 8–11 (finished grade school), 12 (graduated high school), 13–15 (some college), 16 (graduated college), and 17 or more (some postgraduate education). Such a grouping captures the most meaningful distinctions in the educational distribution and preserves the information that would be important for many analyses (see Exhibit 9.12).

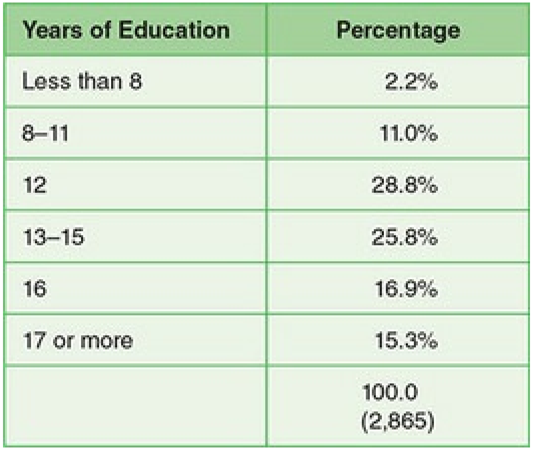
Exhibit 9.11 Grouped Versus Ungrouped Frequency Distributions

Ungrouped		Grouped	
Age	Percentage	Age	Percentage
18	0.2%	18–19	1.4%
19	1.2%	20–29	15.4%
20	0.9%	30–39	17.8%
21	1.2%	40–49	15.5%
22	1.5%	50–59	19.8%
23	1.7%	60–69	16.3%
24	1.2%	70–79	8.6%
25	2.0%	80 or older	5.1%
26	1.5%		100.0% (2,857)
27	2.0%		
28	1.5%		
29	2.0%		
30	1.9%		
31	2.0%		
32	1.5%		
33	1.9%		
34	1.7%		
35	2.0%		
36	1.8%		
37	2.0%		
38	1.5%		
39	1.5%		
40	1.6%		
41	1.3%		
42	1.8%		
43	1.6%		
44	1.8%		
45	0.9%		
46	1.6%		
. . .	. . .		

Source: General Social Survey National Opinion Research Center 2016.

It is also easy to imagine how the requirement that categories be mutually exclusive can be violated. You sometimes see frequency distributions or categories in questionnaires that use such overlapping age categories as 20–30, 30–40, and so on instead of mutually exclusive categories such as those in Exhibit 9.11. The problem is that we then can't tell which category to place someone in who is age 30, 40, and so on.

Exhibit 9.12 Years of Education Completed



Years of Education	Percentage
Less than 8	2.2%
8–11	11.0%
12	28.8%
13–15	25.8%
16	16.9%
17 or more	15.3%
	100.0 (2,865)

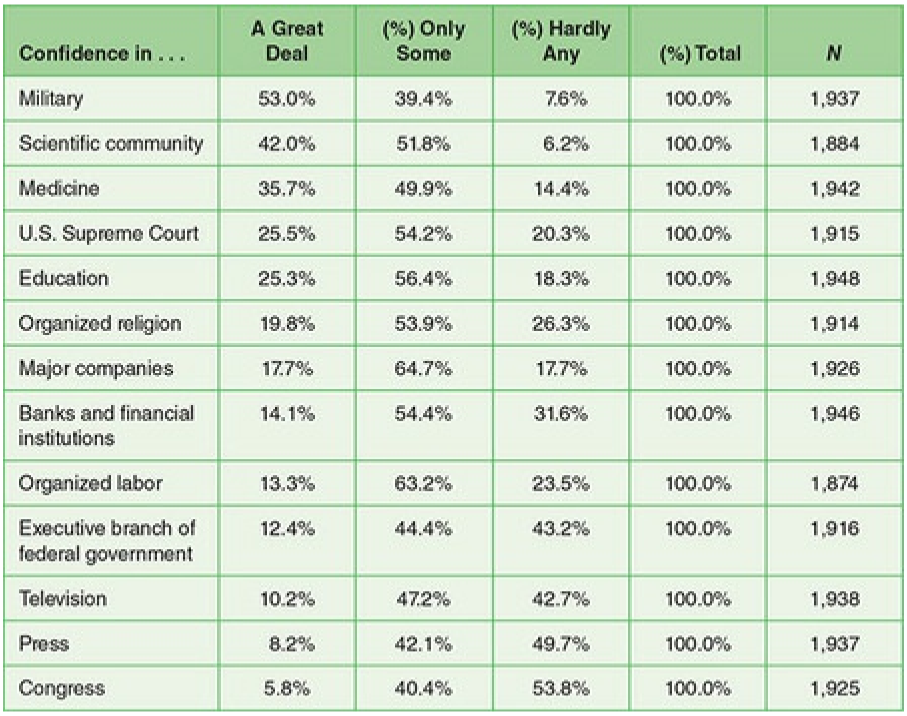
Source: General Social Survey National Opinion Research Center 2016.

Combined and Compressed Distributions

In a combined frequency display, the distributions for a set of conceptually similar variables having the same response categories are presented together. Exhibit 9.13 is a combined display reporting the frequency distributions in percentage form for 13 variables that indicate GSS respondents' level of confidence in U.S. institutions. The different variables are identified in the leftmost column, and their values are labeled along the top. By looking at the table, you can see quickly that confidence is greatest in the military, the scientific community, and medicine; educational institutions, the Supreme Court, and organized religion are regarded with somewhat less confidence. Smaller portions of the U.S. public have much confidence in major companies, banks and financial institutions, organized labor, and the executive branch. Television, the press, and the U.S. Congress elicit the least confidence. Note that the specific variables are ordered in decreasing order of greatest confidence to make it easier to see the pattern. The number of cases on which the distributions are based is included for each variable.

Combined frequency display: A table that presents together the distributions for a set of conceptually similar variables having the same response categories; common headings are used for the responses.

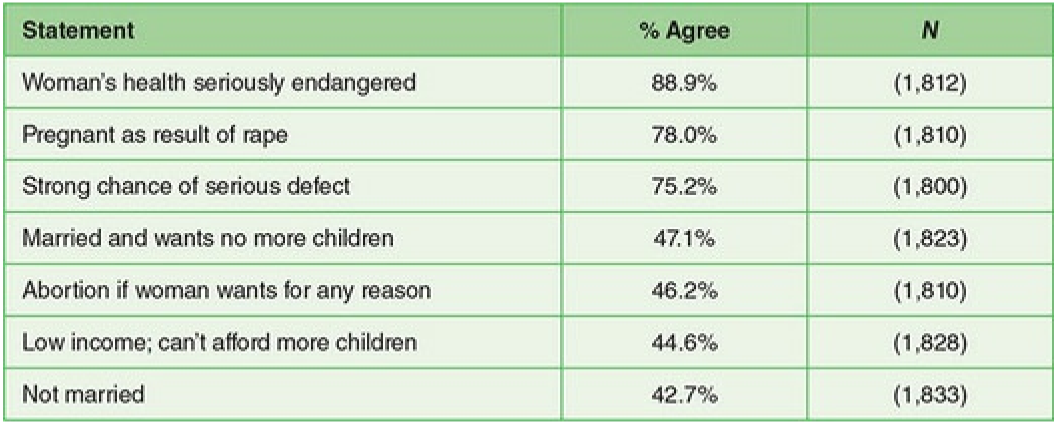
Exhibit 9.13 Confidence in Institutions



Confidence in . . .	A Great Deal	(%) Only Some	(%) Hardly Any	(%) Total	*N*
Military	53.0%	39.4%	7.6%	100.0%	1,937
Scientific community	42.0%	51.8%	6.2%	100.0%	1,884
Medicine	35.7%	49.9%	14.4%	100.0%	1,942
U.S. Supreme Court	25.5%	54.2%	20.3%	100.0%	1,915
Education	25.3%	56.4%	18.3%	100.0%	1,948
Organized religion	19.8%	53.9%	26.3%	100.0%	1,914
Major companies	17.7%	64.7%	17.7%	100.0%	1,926
Banks and financial institutions	14.1%	54.4%	31.6%	100.0%	1,946
Organized labor	13.3%	63.2%	23.5%	100.0%	1,874
Executive branch of federal government	12.4%	44.4%	43.2%	100.0%	1,916
Television	10.2%	47.2%	42.7%	100.0%	1,938
Press	8.2%	42.1%	49.7%	100.0%	1,937
Congress	5.8%	40.4%	53.8%	100.0%	1,925

Source: General Social Survey National Opinion Research Center 2016.

Exhibit 9.14 Agreement With Allowing Abortion, Given Different Conditions



Statement	% Agree	*N*
Woman's health seriously endangered	88.9%	(1,812)
Pregnant as result of rape	78.0%	(1,810)
Strong chance of serious defect	75.2%	(1,800)
Married and wants no more children	47.1%	(1,823)
Abortion if woman wants for any reason	46.2%	(1,810)
Low income; can't afford more children	44.6%	(1,828)
Not married	42.7%	(1,833)

Source: General Social Survey National Opinion Research Center 2016.

Compressed frequency displays can also be used to present cross-tabular data and summary statistics more efficiently, by eliminating unnecessary percentages (such as those corresponding to the second value of a dichotomous variable) and by reducing the need for repetitive labels. Exhibit 9.14 presents a compressed display of agreement that abortion should be allowed given particular conditions. Note that this display presents (in parentheses) the number of cases on which the percentages are based.

Combined and compressed statistical displays facilitate the presentation of a large amount of data in a relatively small space. They should be used with caution, however, because they may baffle people who are not used to them.

Compressed frequency display: A table that presents cross-classification data efficiently by eliminating unnecessary percentages, such as the percentage corresponding to the second value of a dichotomous variable.

Summarizing Univariate Distributions

Summary statistics focus attention on particular aspects of a distribution and facilitate comparison between distributions. For example, if your purpose were to report variation in income by state in a form that is easy for most audiences to understand, you would usually be better off presenting average incomes; many people would find it difficult to make sense of a display containing 50 frequency distributions, although they could readily comprehend a long list of average incomes. A display of average incomes would also be preferable to multiple frequency distributions if your only purpose were to provide a general idea of income differences between states.

In the News Research in the News: Why Key State Polls Were Wrong About Trump

For Further Thought?

Preelection polls in battleground states failed to predict the winner of the 2016 presidential race in those states. Does this mean political polling can no longer be trusted? Participants in the 2017 conference of the American Association of Public Opinion Research (AAPOR) considered the evidence. Three problems accounted for the mistaken polling forecasts: undecided voters swung to Donald Trump by a considerable margin at the last minute—too late to be detected by the polls; likely Hillary Clinton voters were less likely to turn out and vote; and most polls did not adjust for the tendency of well-educated persons to be more likely to respond to surveys (important in the 2016 election because the college-educated voters were much more likely to prefer Clinton).

The lesson seems to be that the failure of the polls can be understood and that the reasons for it can be corrected in the future. Social reality is often more complex than we realize!

1. How convinced are you by the explanations presented at AAPOR? Can you identify other possible sources of error in such political polling?
2. Is it possible to improve polling accuracy by taking advantage of the widespread use of smartphones and social media? What innovations would you suggest based on your own experience? Could greater reliance on "high tech" introduce new problems?

News source: Cohn, Nate. 2017. "Election Review: Why Crucial State Polls Turned Out to Be Wrong." *The New York Times,* June 1, p. A12

Of course, representing a distribution in one number loses information about other aspects of the distribution's shape and so creates the possibility of obscuring important information. If you need to inform a discussion about differences in income inequality between states, for example, measures of central tendency and variability would miss the point entirely. You would either have to present the 50 frequency distributions or use some special statistics that represent the unevenness of a distribution. For this reason, analysts who report summary measures of central tendency usually also report a summary measure of variability and sometimes several measures of central tendency, variability, or both.

Measures of Central Tendency

Central tendency is usually summarized with one of three statistics: the mode, the median, or the mean. For any particular application, one of these statistics may be preferable, but each has a role to play in data analysis. To choose an appropriate measure of central tendency, the analyst must consider a variable's level of measurement, the skewness of a quantitative variable's distribution, and the purpose for which the statistic is used. In addition, the analyst's personal experiences and preferences inevitably will play a role.

Mode

The mode is the most frequent value in a distribution. It is also termed the probability average because, being the most frequent value, it is the most probable. For example, if you were to pick a case at random from the distribution of political views (refer back to Exhibit 9.10), the probability of the case being a moderate would be .374 out of 1, or 37.4%—the most probable value in the distribution.

Mode: The most frequent value in a distribution; also termed the probability average.

The mode is used much less often than the other two measures of central tendency because it can so easily give a misleading impression of a distribution's central tendency. One problem with the mode occurs when a distribution is bimodal, in contrast to being unimodal. A bimodal (or trimodal, etc.) distribution has two or more categories with an equal number of cases and with more cases than any of the other categories. There is no single mode. Imagine that a particular distribution has two categories, each having just about the same number of cases (and these are the two most frequent categories). Strictly speaking, the mode would be the one with more cases, even though the other frequent category had only slightly fewer cases. Another potential problem with the mode is that it might happen to fall far from the main clustering of cases in a distribution. It would be misleading in most circumstances to say simply that the variable's central tendency was whatever the modal value was.

Nevertheless, there are occasions when the mode is very appropriate. Most important, the mode is the only measure of central tendency that can be used to characterize the central tendency of variables measured at the nominal level. We can't say much more about the central tendency of the distribution of marital status in Exhibit 9.5 than that the most common value is married. The mode also is often referred to in descriptions of the shape of a distribution. The terms *unimodal* and *bimodal* appear frequently, as do descriptive statements such as "The typical [most probable] respondent was in her 30s." Of course, when the issue is what the most probable value is, the mode is the appropriate statistic. Which ethnic group is most common in a given school? The mode provides the answer.

Median

The median is the position average, or the point that divides the distribution in half (the 50th percentile). The median is inappropriate for variables measured at the nominal level because their values cannot be put in order, and so there is no meaningful middle position. To determine the median, we simply array a distribution's values in numerical order and find the value of the case that has an equal number of cases above and below it. If the median point falls between two cases (which happens if the distribution has an even number of cases), the median is defined as the average of the two middle values and is computed by adding the values of the two middle cases and dividing by 2.

The median in a frequency distribution is determined by identifying the value corresponding to a cumulative percentage of 50. Starting at the top of the years of education distribution in Exhibit 9.12, for example, and adding up the percentages, we find that we have reached 42% in the 12 years category and then 67.8% in the 13–15 years category. The median is therefore 13–15.

With most variables, it is preferable to compute the median from ungrouped data because that method results in an exact value for the median, rather than an interval. In the grouped age distribution in Exhibit 9.11, for example, the median is in the 40s interval. But if we determine the median from the ungrouped data, we can state that the exact value of the median is 49.

Mean

The mean, or arithmetic average, considers the values of each case in a distribution—it is a weighted average. The mean is computed by adding up the value of all the cases and dividing by the total number of cases, thereby accounting for the value of each case in the distribution:

Mean = Sum of value of cases/Number of cases

In algebraic notation, the equation is $\bar{Y} = \sum Y_i / N$ For example, to calculate the mean of eight (hypothetical) cases, we add the values of all the cases (ΣY_i) and divide by the number of cases (N):

$$(28 + 117 + 42 + 10 + 77 + 51 + 64 + 55)/8 = 444/8 = 55.5$$

Bimodal: A distribution that has two nonadjacent categories with about the same number of cases, and these categories have more cases than any others.

Unimodal: A distribution of a variable in which there is only one value that is the most frequent.

Median: The position average, or the point that divides a distribution in half (the 50th percentile).

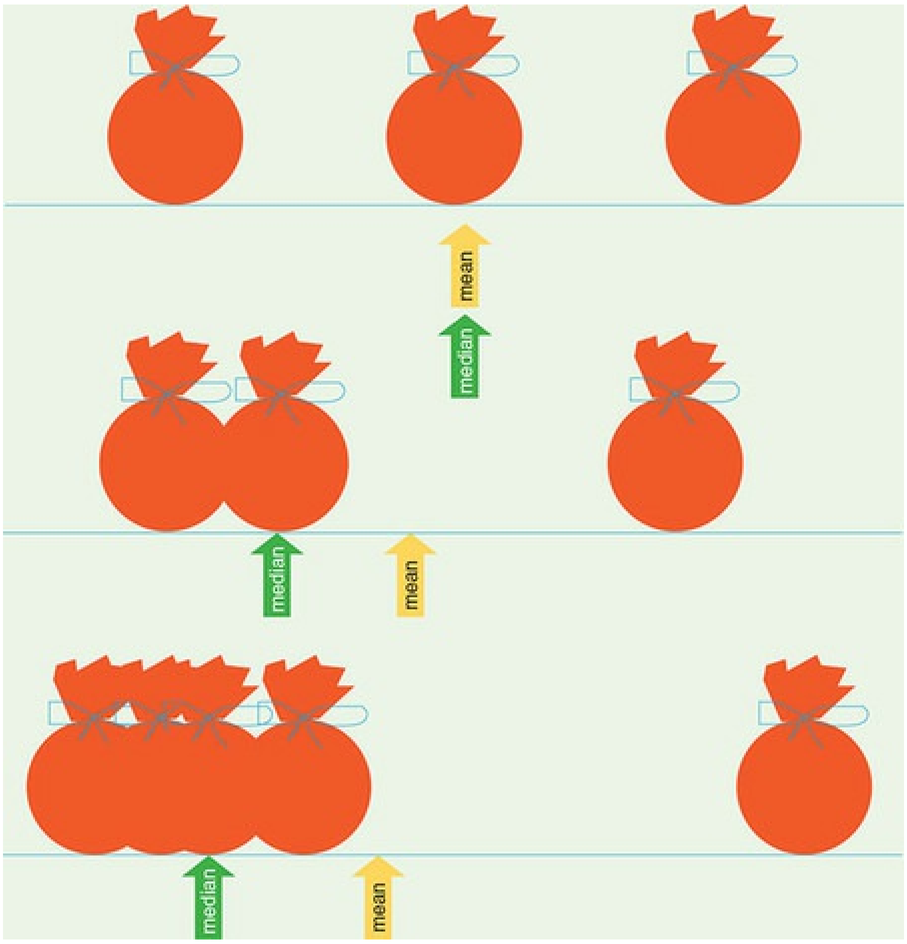
Mean: The arithmetic, or weighted, average, computed by adding the value of all the cases and dividing by the total number of cases.

Because computing the mean requires adding the values of the cases, it makes sense to compute a mean only if the values of the cases can be treated as actual quantities—that is, if they reflect an interval or ratio level of measurement, or if they are ordinal and we assume that ordinal measures can be treated as interval. It would make no sense to calculate the mean religion. For example, imagine a group of four people in which there were two Protestants, one Catholic, and one Jew. To calculate the mean, you would need to solve the equation (Protestant + Protestant + Catholic + Jew) + 4 = ? Even if you decide that Protestant = 1, Catholic = 2, and Jew = 3 for data entry purposes, it still doesn't make sense to add these numbers because they don't represent quantities of religion.

Median or Mean?

Both the median and the mean are used to summarize the central tendency of quantitative variables, but their suitability for a particular application must be carefully assessed. The key issues to be considered in this assessment are the variable's level of measurement, the shape of its distribution, and the purpose of the statistical summary. Consideration of these issues will sometimes result in a decision to use both the median and the mean and will sometimes result in neither measure being seen as preferable. But in many other situations, the choice between the mean and median will be clear-cut as soon as the researcher takes the time to consider these three issues.

Exhibit 9.15 The Mean as a Balance Point



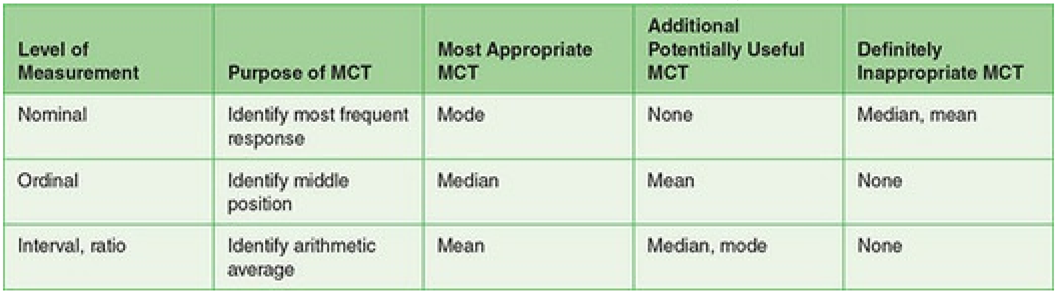
Level of measurement is a key concern because to calculate the mean, we must add the values of all the cases—a procedure that assumes the variable is measured at the interval or ratio level. So even though we know that coding *Agree* as 2 and *Disagree* as 3 does not really mean that *Disagree* is 1 unit more of disagreement than *Agree,* the mean assumes this evaluation to be true. Because calculation of the median requires only that we order the values of cases, we do not have to make this assumption. Technically speaking, then, the mean is simply an inappropriate statistic for variables measured at the ordinal level (and you already know that it is completely meaningless for variables measured at the nominal level). In practice, however, many social researchers use the mean to describe the central tendency of variables measured at the ordinal level, for the reasons outlined earlier.

The shape of a variable's distribution should also be considered when deciding whether to use the median or mean. When a distribution is perfectly symmetric, so that the distribution of values below the median is a mirror image of the distribution of values above the median, the mean and median will be the same. But the values of the mean and median are affected differently by skewness, or the presence of cases with extreme values on one side of the distribution but not the other side. Because the median accounts for only the number of cases above and below the median point, not the value of these cases, it is not affected in any way by extreme values. Because the mean is based on adding the value of all the cases, it will be pulled in the direction of exceptionally high (or low) values. When the value of the mean is larger than the value of the median, we know that the distribution is skewed in a positive direction, with proportionately more cases with higher than lower values. When the mean is smaller than the median, the distribution is skewed in a negative direction.

This differential impact of skewness on the median and mean is illustrated in Exhibit 9.15. On the first balance beam, the cases (bags) are spread out equally, and the median and mean are in the same location. On the second and third balance beams, the median corresponds to the value of the middle case, but the mean is pulled toward the value of the one case with an extremely high value. For this reason, the mean age (49.2) for the 2,857 cases represented partially in the detailed age distribution in Exhibit 9.11 is slightly higher than the median age (49.0). Although in the distribution represented in Exhibit 9.11 the difference is small, in some distributions, the two measures will have markedly different values, and in such instances, the median may be preferred.

The single most important influence on the choice of the median or the mean for summarizing the central tendency of quantitative variables should be the purpose of the statistical summary. If the purpose is to report the middle position in one or more distributions, then the median is the appropriate statistic, whether or not the distribution is skewed (see Exhibit 9.16). For example, with respect to the age distribution from the GSS, you could report that half the U.S. population is younger than 49 years old, and half the population is older than that. But if the purpose is to show how likely different groups are to have age-related health problems, then the measure of central tendency for these groups should account for people's actual ages, not just the number of people who are older and younger than a particular age. For this purpose, the median would be inappropriate because it would not distinguish between two distributions that have the same median but with different numbers of older people. In one distribution, everyone might be between the ages of 35 and 55, with a median of 49. In another distribution with a median of 49, half of the cases could have ages above 60. The mean of the second distribution would be higher, reflecting the fact that it has a greater number of older people.

Exhibit 9.16 Selection of Measures of Central Tendency (MCT)



Level of Measurement	Purpose of MCT	Most Appropriate MCT	Additional Potentially Useful MCT	Definitely Inappropriate MCT
Nominal	Identify most frequent response	Mode	None	Median, mean
Ordinal	Identify middle position	Median	Mean	None
Interval, ratio	Identify arithmetic average	Mean	Median, mode	None

Keep in mind that it is not appropriate to use either the median or the mean as a measure of central tendency for variables measured at the nominal level because at this level the different attributes of a variable cannot be ordered as higher or lower (as reflected in Exhibit

9.16). Technically speaking, the mode should be used to measure the central tendency of variables measured at the nominal level (and it can also be used with variables measured at the ordinal, interval, and ratio levels). The median is most suited to measure the central tendency of variables measured at the ordinal level (and it can also be used to measure the central tendency of variables measured at the interval and ratio levels). Finally, the mean is only unequivocally suited to measure the central tendency for variables measured at the interval and ratio levels.

It is not entirely legitimate to represent the central tendency of a variable measured at the ordinal level with the mean: Calculation of the mean requires summing the values of all cases, and at the ordinal level, these values indicate only order, not actual numbers. Nonetheless, many social scientists use the mean with ordinal-level variables and find that this is potentially useful for comparisons between variables and as a first step in more complex statistical analyses. The median and mode can also be useful as measures of central tendency for variables measured at the interval and ratio levels, when the goal is to indicate middle position (the median) or the most frequent value (the mode).

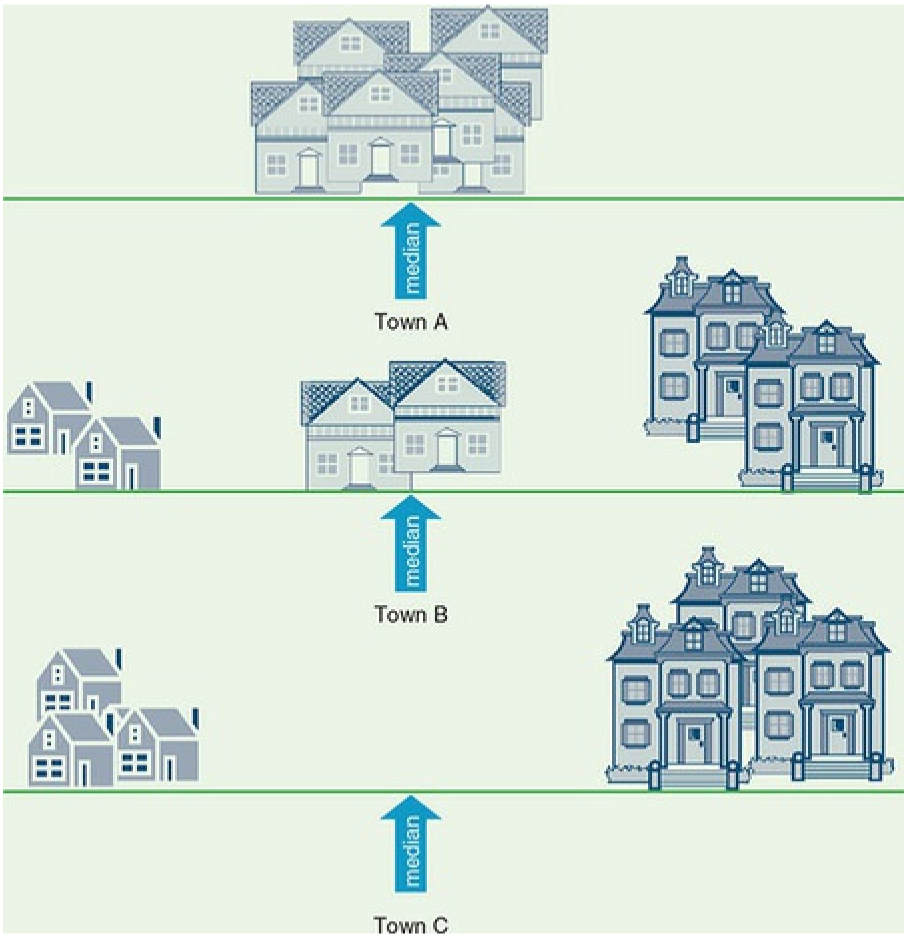
In general, the mean is the most commonly used measure of central tendency for quantitative variables, both because it accounts for the value of all cases in the distribution and because it is the foundation for many more advanced statistics. However, the mean's very popularity results in its use in situations for which it is inappropriate. Keep an eye out for this problem.

Measures of Variation

You already have learned that central tendency is only one aspect of the shape of a distribution—the most important aspect for many purposes but still just a piece of the total picture. A summary of distributions based only on their central tendency can be incomplete, even misleading. For example, three towns might have the same median income but still be very different in their social character because of the shape of their income distributions. As illustrated in Exhibit 9.17, Town A is a homogeneous middle-class community; Town B is very heterogeneous; and Town C has a polarized, bimodal income distribution, with mostly very poor and very rich people and few in between. However, all three towns have the same median income.

The way to capture these differences is with statistical measures of variation. Four popular measures of variation are the range, the interquartile range, the variance, and the standard deviation (which is the most popular measure of variability). To calculate each of these measures, the variable must be at the interval or ratio level (but many would argue that, like the mean, they can be used with ordinal-level measures, too). Statistical measures of variation are used infrequently with variables measured at the nominal level, so these measures will not be presented here.

Exhibit 9.17 Distributions Differing in Variability but Not Central Tendency



It’s important to realize that measures of variability are summary statistics that capture only part of what we need to be concerned with about the distribution of a variable. In particular, they do not tell us about the extent to which a distribution is skewed, which we’ve seen is very important for interpreting measures of central tendency. Researchers usually evaluate the skewness of distributions just by eyeballing them.

Range

The range is a simple measure of variation, calculated as the highest value in a distribution minus the lowest value:

Range = Highest value – Lowest value

Range: The true upper limit in a distribution minus the true lower limit (or the highest rounded value minus the lowest rounded value, plus one).

It often is important to report the range of a distribution to identify the whole range of possible values that might be encountered. However, because the range can be drastically altered by just one exceptionally high or low value (termed an outlier), it does not do an adequate job of summarizing the extent of variability in a distribution.

Interquartile Range

A version of the range statistic, the interquartile range, avoids the problem created by outliers. Quartiles are the points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases. You already know how to determine the second quartile, corresponding to the point in the distribution covering half of the cases—it is another name for the median. The first and third quartiles are determined in the same way but by finding the points corresponding to 25% and 75% of the cases, respectively. The interquartile range is the difference between the end of the first quartile and the beginning of the third quartile.

We can use the distribution of age for an example. If you add up the percentages corresponding to each value of age (ungrouped) in Exhibit 9.11, you'll find that you reach the first quartile (25% of the cases) at the age value of 34. If you were to continue, you would find that age 62 corresponds to the third quartile—the point where you have covered 75% of the cases. So the interquartile range for age, in the GSS 2016 data, is 28:

Third quartile − First quartile = Interquartile range

$62 - 34 = 28$

Variance

The variance is the average squared deviation of each case from the mean, so it accounts for the amount by which each case differs from the mean. An example of how to calculate the variance, using the following formula, appears in Exhibit 9.18:
$\sigma 2 = \sum (Y i - \bar{Y} i) N$

Symbol key: $\bar{Y}$ = mean; N = number of cases; Σ = sum over all cases; Y_i = value of variable Y for case *i;* s^2 = variance.

You can see in Exhibit 9.18 two examples of summing over all cases, the operation represented by the Greek letter Σ in the formula ($\bar{Y}$= 24.27).

The variance is used in many other statistics, although it is more conventional to measure variability with the closely related standard deviation than with the variance.

Standard Deviation

The standard deviation is simply the square root of the variance. It is the square root of the average squared deviation of each case from the mean:
$\sigma 2 = \sum (Y i - Y ^{-} i) 2 N$

Outlier: An exceptionally high or low value in a distribution.

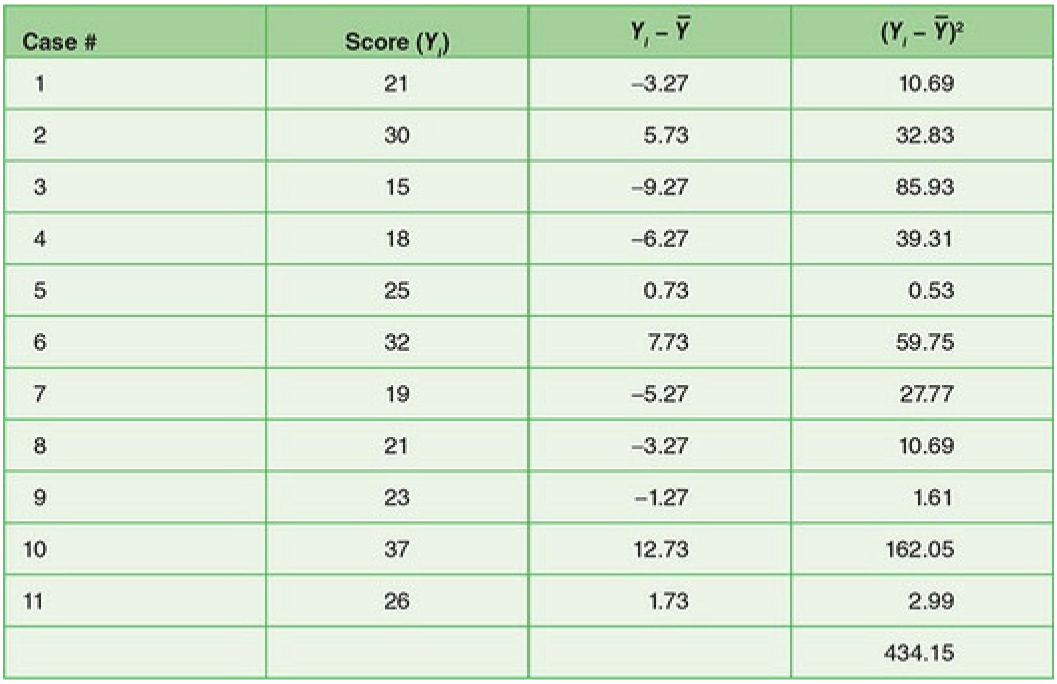
Interquartile range: The range in a distribution between the end of the first quartile and the beginning of the third quartile.

Quartiles: The points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases.

Variance: A statistic that measures the variability of a distribution as the average squared deviation of each case from the mean.

Standard deviation: The square root of the average squared deviation of each case from the mean.

Exhibit 9.18 Calculation of the Variance



Case #	Score (Y_i)	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	21	–3.27	10.69
2	30	5.73	32.83
3	15	–9.27	85.93
4	18	–6.27	39.31
5	25	0.73	0.53
6	32	7.73	59.75
7	19	–5.27	27.77
8	21	–3.27	10.69
9	23	–1.27	1.61
10	37	12.73	162.05
11	26	1.73	2.99
			434.15

Symbol key: $\bar{Y}$ = mean; N = number of cases; Σ = sum over all cases; Y_i = value of variable Y for case *i;* √= square root; s = standard deviation.

When the standard deviation is calculated from sample data, the denominator is supposed to be $N - 1$, rather than N, an adjustment that has no discernible effect when the number of cases is reasonably large. You also should note that the use of *squared* deviations in the formula accentuates the impact of relatively large deviations because squaring a large number makes that number count much more.

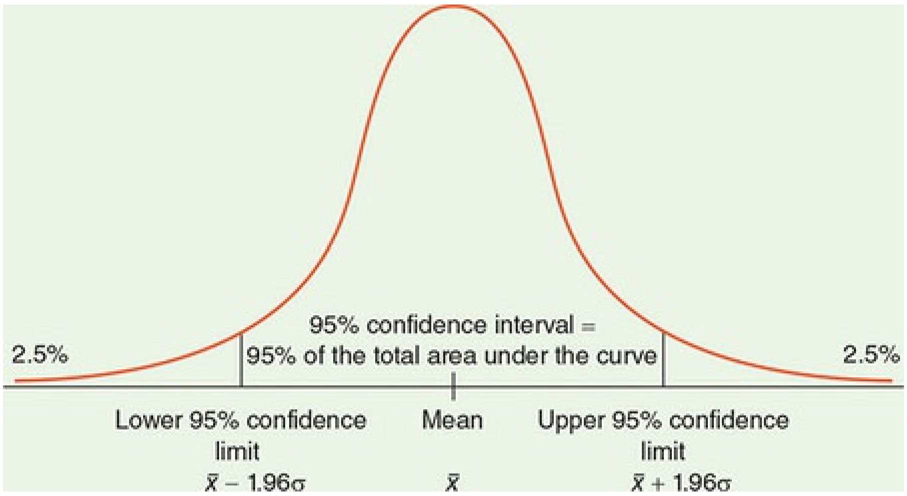
The standard deviation has mathematical properties that increase its value for statisticians. You already learned about the normal distribution in Chapter 5. A normal distribution is a distribution that results from chance variation around a mean. It is symmetric and tapers off in a characteristic shape from its mean. If a variable is normally distributed, 68% of the cases will lie between plus and minus 1 standard deviation from the distribution's mean, and 95% of the cases will lie between plus and minus 1.96 standard deviations from the mean (see Exhibit 9.19).

Normal distribution: A symmetric, bell-shaped distribution that results from chance variation around a central value.

This correspondence of the standard deviation to the normal distribution enables us to infer how confident we can be that the mean (or some other statistic) of a population sampled randomly is within a certain range of the sample mean. This is the logic behind calculating confidence limits around the mean. You learned in Chapter 5 that confidence limits indicate how confident we can be, based on our random sample, that the value of some statistic in the population falls within a particular range. (The actual value in the population is the *population parameter.*) Now that you know how to compute the standard deviation, it is just a short additional step to computation of the confidence limits around a mean. There are just four more steps:

1. Calculate the standard error. This is the estimated value of the standard deviation of the sampling distribution from which your sample was selected. SE= $\sigma/(n-1)$. In words, divide the standard of the sample by the square root of the number of cases in the sample minus one.

Exhibit 9.19 The Normal Distribution



2. Decide on the degree of confidence that you want to have that the population parameter falls within the confidence interval you compute. It is conventional to calculate the 95%, 99%, or even the 99.9% confidence limits around the mean. Most often, the 95% confidence limits are used, so I will show the calculation for this estimate.
3. Multiply the value of the SE ´ 1.96. This is because 95% of the area under the normal curve falls within ±1.96 standard deviation units of the mean.
4. Add and subtract the number calculated in (3) from the sample mean. The resulting numbers are the upper and lower confidence limits.

If you had conducted these steps for age with the 2016 GSS data, you would now be able to report, "Based on the GSS 2016 sample, I can be 95% confident that the true mean age in the population is between 48.51 and 49.81" (49.16 - .331*1.96 and 49.16 + .331*1.96). When you read in media reports about polling results that the "margin of error" was ±3 points, you'll now know that the pollster was simply providing 95% confidence limits for the statistic.

Analyzing Data Ethically: How Not to Lie With Statistics

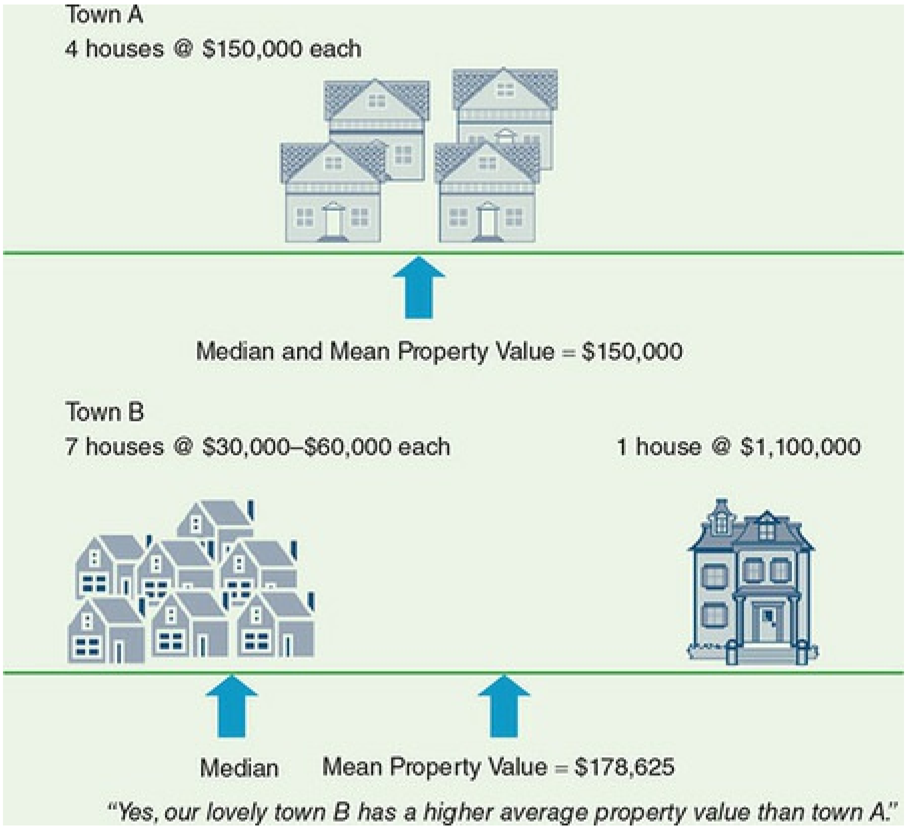
Using statistics ethically means first and foremost being honest and open. Findings should be reported honestly, and the researcher should be open about the thinking that guided his or her decision to use particular statistics. Although this section has a humorous title (after Darrell Huff's [1954] little classic, *How to Lie With Statistics*), make no mistake about the intent: It is possible to distort social reality with statistics, and it is unethical to do so knowingly, even when the error is due more to carelessness than deceptive intent.

Summary statistics can easily be used unethically, knowingly or not. When we summarize a distribution in a single number, even in two numbers, we are losing much information. Neither central tendency nor variation describes a distribution's overall shape. And taken separately, neither measure tells us about the other characteristic of the distribution (central tendency or variation). So reports using measures of central tendency should normally also include measures of variation. And we should inspect the shape of any distribution for which we report summary statistics to ensure that the summary statistic does not mislead us (or anyone else) because of an unusual degree of skewness.

It is possible to mislead those who read statistical reports by choosing summary statistics that accentuate a particular feature of a distribution. For example, imagine an unscrupulous realtor trying to convince a prospective home buyer in Community B that it is a community with very high property values, when it actually has a positively skewed distribution of property values (see Exhibit 9.20). The realtor compares the mean price of homes in Community B to that for Community A (one with a homogeneous mid-priced set of homes) and therefore makes Community B look much better. In truth, the higher mean in Community B reflects a very skewed, lopsided distribution of property values—most residents own small, cheap homes. A median would provide a better basis for comparison.

You have already seen that it is possible to distort the shape of a distribution by ignoring some of the guidelines for constructing graphs and frequency distributions. Whenever you need to group data in a frequency distribution or graph, you can reduce the potential for problems by inspecting the ungrouped distributions and then using a grouping procedure that does not distort the distribution's basic shape. When you create graphs, be sure to consider how the axes you choose may change the distribution's apparent shape.

Exhibit 9.20 Using the Mean to Create a More Favorable Impression


Town A
4 houses @ $150,000 each
Median and Mean Property Value = $150,000
Town B
7 houses @ $30,000–$60,000 each
1 house @ $1,100,000
Median
Mean Property Value = $178,625
"Yes, our lovely town B has a higher average property value than town A."

Cross-Tabulating Variables

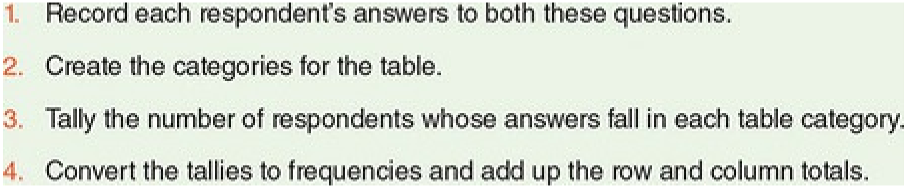
Most data analyses focus on relationships between variables to test hypotheses or just to describe or explore relationships. For each of these purposes, we must examine the association between two or more variables. Cross-tabulation (crosstab) is one of the simplest methods for doing so. A cross-tabulation, or contingency table, displays the distribution of one variable for each category of another variable; it can also be termed a *bivariate distribution.* You can also display the association between two variables in a graph; we will see an example in this section. In addition, crosstabs provide a simple tool for statistically controlling one or more variables while examining the associations between others. In the next section, you will learn how crosstabs used in this way can help test for spurious relationships and evaluate causal models. We will examine several trivariate tables.

Cross-tabulation (crosstab) or contingency table: In the simplest case, a bivariate (two-variable) distribution, showing the distribution of one variable for each category of another variable; can be elaborated using three or more variables.

Constructing Contingency Tables

The exhibits throughout this section are based on 2016 GSS data. You can learn in the SPSS exercises at the end of this chapter and online how to use the SPSS program to generate cross-tabulations from a data set that has been prepared for analysis with SPSS. But let's now briefly see how you would generate a cross-tabulation without a computer. Exhibit 9.21 shows the basic procedure, based on results of responses in a survey about income and voting:

Exhibit 9.21 Converting Data From Case Records (Raw Data) Into a Crosstab



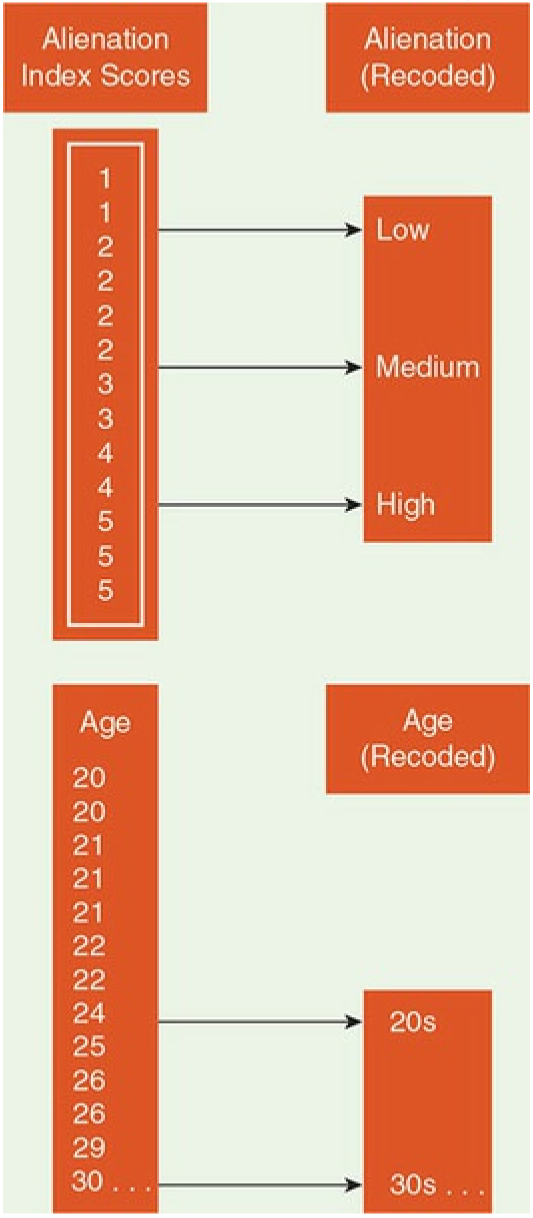
1. Record each respondent's answers to both these questions.
2. Create the categories for the table.
3. Tally the number of respondents whose answers fall in each table category.
4. Convert the tallies to frequencies and add up the row and column totals.

Simple, isn't it? Now, you're ready to shift to inspecting crosstabs produced with the SPSS computer program for the 2016 GSS data. We'll consider in just a bit the procedures for converting frequencies in a table into percentages and for interpreting, or "reading," crosstabs.

Cross-tabulation is a useful method for examining the relationship between variables only when they have just a few categories. For most analyses, 10 categories is a reasonable upper limit, but even 10 is too many unless you have a pretty large number of cases (more than 100). If you wish to include in a crosstab a variable with many categories, or one that varies along a continuum with many values, you should first recode the values of that variable to a smaller number. For example, you might recode the values of a 5-point index to just *high (representing scores of 4 and 5), medium (3),* and *low (1 and 2).* You might recode the numerical values of age to 10-year intervals. Exhibit 9.22 provides examples.

Exhibit 9.23 displays the cross-tabulation of voting by family income, using the 2016 GSS data, so that we can test the hypothesis that likelihood of voting increases with this one social status indicator. The table is presented first with frequencies and then again with percentages. In both tables, the *body* of the table is the part between the row and column labels and the row and column totals. The cells of the table are defined by combinations of row and column values. Each cell represents cases with a unique combination of values of the two variables, corresponding to that particular row and column. The marginal distributions of the table are on the right (the *row marginals*) and underneath (the *column marginals*). These are just the frequency distributions for the two variables (in number of cases, percentages, or both), considered separately. (The column marginals in Exhibit 9.23 are for family income; the row marginals are for the distribution of voting.) The independent variable is usually the column variable; the dependent variable then is the row variable.

Exhibit 9.22 Examples of Recoding


Alienation Index Scores
1
1
2
2
2
2
3
3
4
4
5
5
5
Alienation (Recoded)
Low
Medium
High
Age
20
20
21
21
21
22
22
24
25
26
26
29
30 . . .
Age (Recoded)
20s
30s . . .

The table in the upper panel of Exhibit 9.23 shows the number of cases with each combination of values of voting and family income. In it, you can see that 357 of those earning less than $25,000 per year voted, whereas 270 did not. On the other hand, 634 of those whose family earned $75,000 or more voted, whereas 148 of these high-income respondents did not vote. It often is hard to look at a table in this form, with just the numbers of cases in each cell, and determine whether there is a relationship between the two variables. We need to convert the cell frequencies into percentages, as in the table in the lower panel of Exhibit 9.23. This table presents the data as percentages within the categories of the independent variable (the column variable, in this case). In other words, the cell frequencies have been converted into percentages of the column totals (the *n* in each column). For example, in Exhibit 9.23 the number of people in families earning less than $25,000 who voted is 357 out of 627, or 56.9%. Because the cell frequencies have been converted to percentages of the column totals, the numbers add up to 100 in each column, but not across the rows.

To read the percentage table (the bottom panel of Exhibit 9.23), compare the percentage distribution of voting across the columns, starting with the lowest income category (in the left column) and moving from left to right. You see that as income increases, the percentage who voted also rises, from 57% (rounding off) of those with annual family incomes less than $25,000 (in the first cell in the first column), to 66% of those with family incomes between $25,000 and $49,999, then 70% of those with family incomes from $50,000 to $74,999, and then up to 81% of those with family incomes of $75,000 or more (the last cell in the body of the table in the first row). This result is consistent with the hypothesis.

When a table is converted to percentages, usually just the percentage in each cell should be presented, not the number of cases in each cell. Include 100% at the bottom of each column (if the independent variable is the column variable) to indicate that the percentages add up to 100, as well as the *base number* (*n*) for each column (in parentheses). If the percentages add up to 99 or 101 because of rounding error, just indicate this in a footnote.

Follow these rules when you create and then read a percentage table:

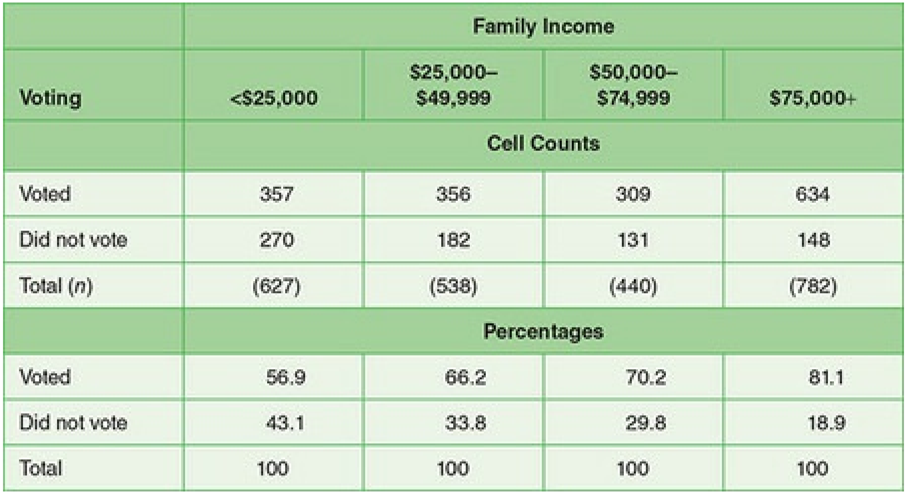
1. Make the independent variable the column variable and the dependent variable the row variable.
2. Percentage the table column by column, on the column totals. The percentages should add to 100 (or perhaps 99 or 101, if there has been rounding error) in each column.
3. Compare the distributions of the dependent variable (the row variable) across each column.

Marginal distribution: The summary distributions in the margins of a cross-tabulation that correspond to the frequency distribution of the row variable and of the column variable.

Percentages: Relative frequencies, computed by dividing the frequency of cases in a particular category by the total number of cases and then multiplying by 100.

Skip ahead to the bottom panel of Exhibit 9.34 and try your hand at the table reading process (as described by Rule 3) with this larger table. The table in the bottom panel describes the association between education and family income. Examine the distribution of family income for those who did not finish high school (first column). More than half (53.4%) reported a family income under $25,000, whereas just 11.6% reported a family income of $75,000 or more. Then, examine the distribution of family income for the respondents who had finished high school but had gone no further. Here, the distribution of family income has shifted upward, with just 32.9% reporting a family income under $25,000 and 20.5% reporting family incomes of $75,000 or more—that's almost three times the percentage in that category than we saw for those who did not finish high school. You can see there are also more respondents in the $50,000 to $74,999 category than there were for those who had not completed high school. Now, examine the column representing those who had completed some college. The percentage with family incomes under $25,000 has dropped again, to 28.4%, whereas the percentage in the highest income category has risen to 27.2%. The college graduates have a much higher family income distribution than do those with less education, with just 10.2% reporting family incomes of less than $25,000 and more than half (55.1%) reporting family incomes of $75,000 or more. If you step back and compare the income distributions across the four categories of education, you see that incomes increase markedly and consistently. The relationship is positive (fortunately, for students like you who are working so hard to finish college!).

Exhibit 9.23 Cross-Tabulation of Voting in 2012 by Family Income: Cell Counts and Percentages

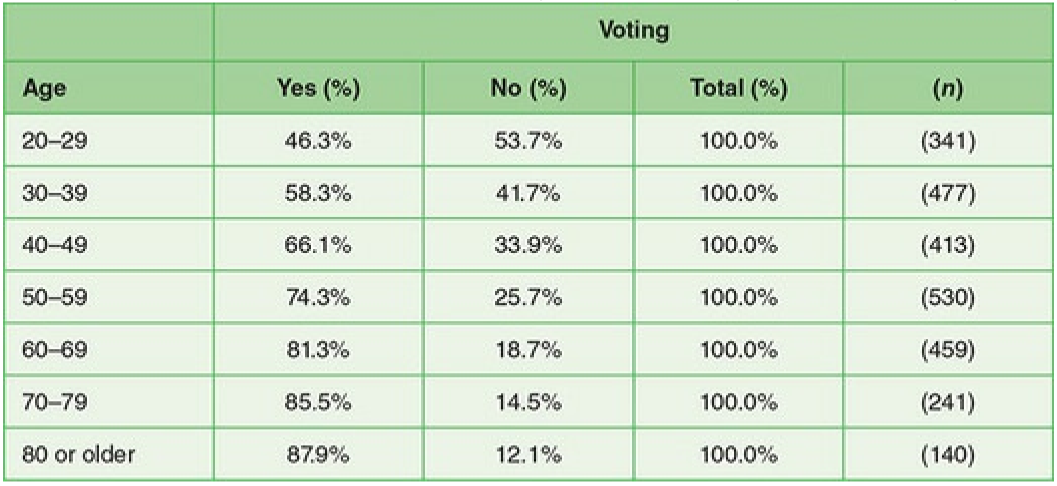


	Family Income			
Voting	**<$25,000**	**$25,000–$49,999**	**$50,000–$74,999**	**$75,000+**
	Cell Counts			
Voted	357	356	309	634
Did not vote	270	182	131	148
Total (*n*)	(627)	(538)	(440)	(782)
	Percentages			
Voted	56.9	66.2	70.2	81.1
Did not vote	43.1	33.8	29.8	18.9
Total	100	100	100	100

Source: General Social Survey National Opinion Research Center 2016.

But the independent variable does not *have* to be the column variable; what is critical is to be consistent within a report or paper. You will find in published articles and research reports some percentage tables in which independent variable and dependent variable positions are reversed. If the independent variable is the row variable, we percentage the table on the row totals (the *n* in each row), and so the percentages total 100 across the rows. Let's examine Exhibit 9.24, which is percentaged on the row variable: age. When you read the table in Exhibit 9.24, you find that 46.3% of those in their 20s voted (and 53.7% didn't vote), compared with 58.3% of those in their 30s, 66.1% and 74.3% of those in their 40s and 50s, respectively, and 81.3% to 87.9% of those ages 60 or older.

Exhibit 9.24 Cross-Tabulation of Voting in 2012 by Age (row percentages)



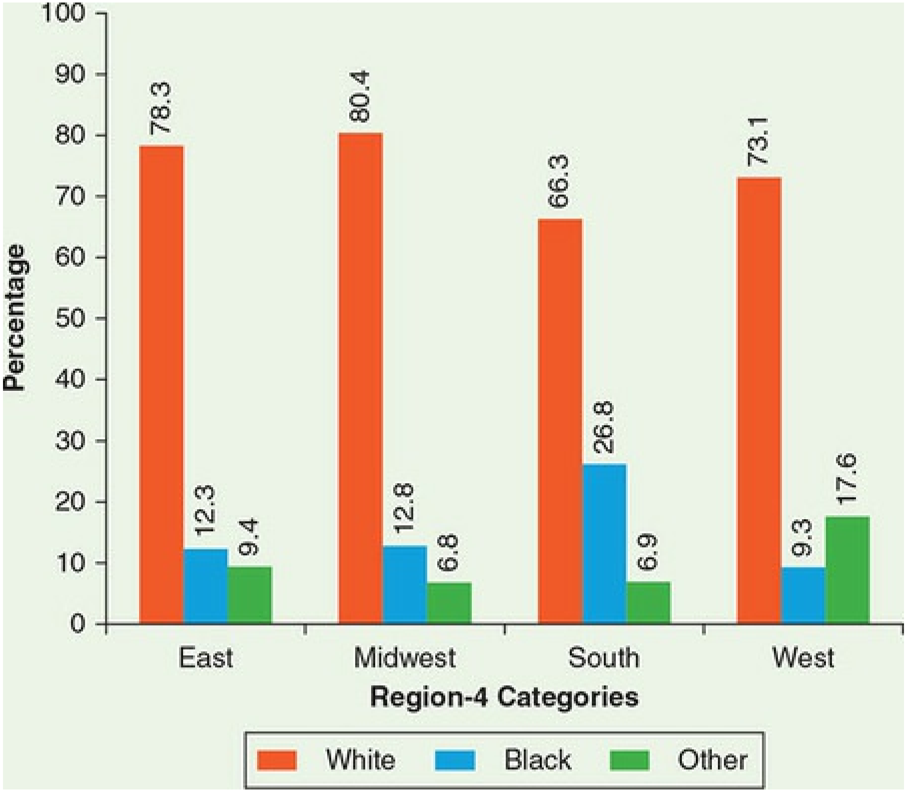
	Voting			
Age	**Yes (%)**	**No (%)**	**Total (%)**	**(*n*)**
20–29	46.3%	53.7%	100.0%	(341)
30–39	58.3%	41.7%	100.0%	(477)
40–49	66.1%	33.9%	100.0%	(413)
50–59	74.3%	25.7%	100.0%	(530)
60–69	81.3%	18.7%	100.0%	(459)
70–79	85.5%	14.5%	100.0%	(241)
80 or older	87.9%	12.1%	100.0%	(140)

Source: General Social Survey National Opinion Research Center 2016.

Graphing Association

Graphs provide an efficient tool for summarizing relationships between variables. Exhibit 9.25 displays the relationship between race and region in graphic form. It shows that the percentage of the population that is black is highest in the South (26.8%), whereas persons of "other" races are most common in the West (17.6%).

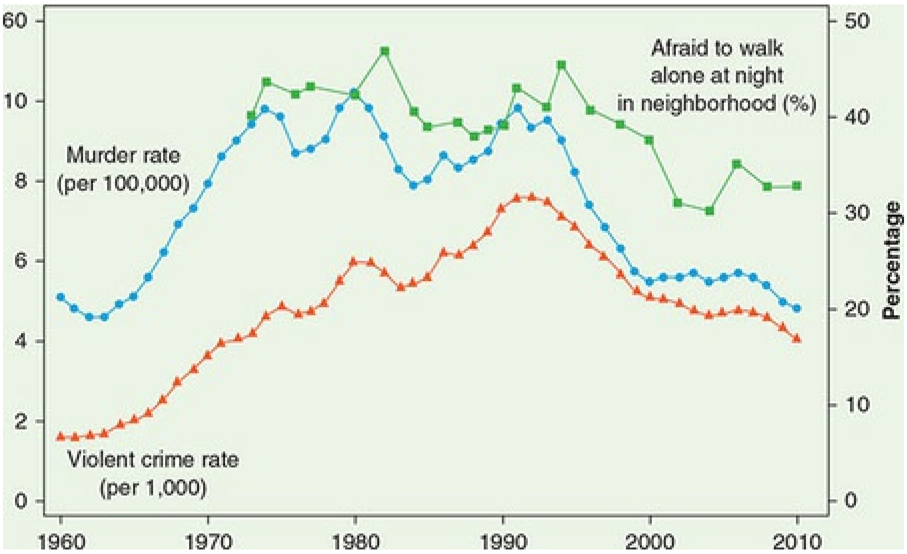
Exhibit 9.25 Race by Region of the United States



Source: General Social Survey National Opinion Research Center 2016.

Another good example of the use of graphs to show relationships is provided by a graph that combines data from the FBI Uniform Crime Reports and the General Social Survey (Egan 2012). Exhibit 9.26 shows how the rate of violent crime, the murder rate, and fear of walking alone at night in the neighborhood have varied over time: The violent crime and murder rates rose in the 1960s (the GSS question was first asked in 1972) and the violent crime rate continued to rise into the early 1990s, although the murder rate—its much less common component—was more variable. Then both crime rates fell through 2010. The fear of walking alone remained relatively constant during the 1970s and 1980s, but then dropped early in the 2000s, although it seems to have leveled off by 2005. Because the three rates displayed in this graph are measured at the interval–ratio level, the graph can represent the variation over time with continuous lines.

Exhibit 9.26 Violence in America, 1960–2010



Source: General Social Survey National Opinion Research Center 2016.

Describing Association

A cross-tabulation table reveals four aspects of the association between two variables:

- *Existence.* Do the percentage distributions vary at all between categories of the independent variable?
- *Strength.* How much do the percentage distributions vary between categories of the independent variable? In most analyses, the analyst would not pay much attention to differences of less than 10 percentage points between categories of the independent variable.
- *Direction.* For quantitative variables, do values on the dependent variable tend to increase or decrease with an increase in value on the independent variable?
- *Pattern.* For quantitative variables, are changes in the percentage distribution of the dependent variable fairly regular (simply increasing or decreasing), or do they vary (perhaps increasing, then decreasing, or perhaps gradually increasing, then rapidly increasing)?

Looking back at Exhibit 9.23, an association exists; it is moderately strong (the difference in percentages between those who voted in the first and last column is 24.2 percentage points); and the direction of association between likelihood of voting and family income is positive. The pattern in this table is close to what is termed monotonic. In a monotonic relationship, the value of cases consistently increases (or decreases) on one variable as the value of cases increases on the other variable. The relationship in the table that we will examine in Exhibit 9.34, involving income and education, is also monotonic. By contrast, the relationship charted in Exhibit 9.26 between the murder rate and year (1960–2010) and between the violent crime rate and year are curvilinear. Both rates increase over the years and then decrease.

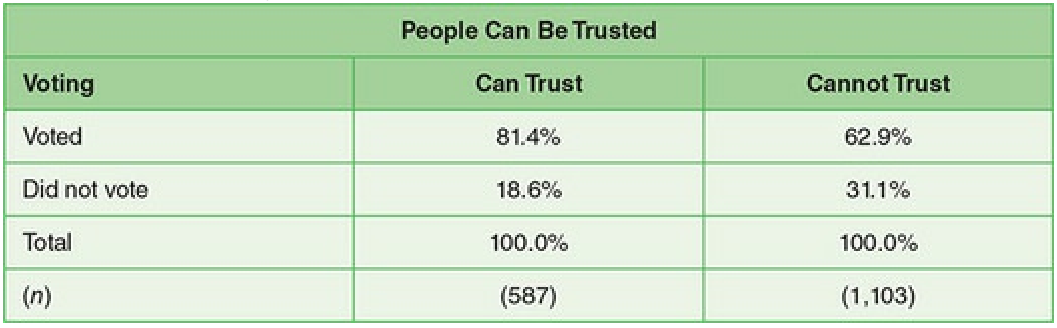
Monotonic: A pattern of association in which the value of cases on one variable increases or decreases fairly regularly across the categories of another variable.

Curvilinear: Any pattern of association between two quantitative variables that does not involve a regular increase or decrease.

The relationship between the measure of trust and voting appears in Exhibit 9.27. There is an association, and in the direction I hypothesized: 81.4% of those who believe that people can be trusted voted, compared with 62.9% of those who believe that people cannot be trusted. Because both variables are dichotomies, there can be no pattern to the association beyond the difference between the two percentages. (Comparing the column percentages in either the first or the second row gives the same picture.)

Exhibit 9.28, by contrast, gives less evidence of an association between gender and voting. The difference between the percentage of men and women who voted is 4 percentage points.

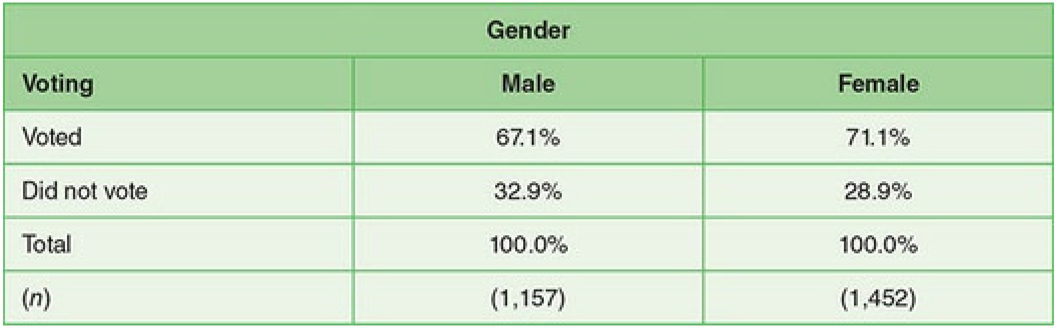
Exhibit 9.27 Voting in 2012 by Interpersonal Trust



People Can Be Trusted		
Voting	**Can Trust**	**Cannot Trust**
Voted	81.4%	62.9%
Did not vote	18.6%	31.1%
Total	100.0%	100.0%
(*n*)	(587)	(1,103)

Source: General Social Survey National Opinion Research Center 2016.

Exhibit 9.28 Voting in 2012 by Gender



Gender		
Voting	**Male**	**Female**
Voted	67.1%	71.1%
Did not vote	32.9%	28.9%
Total	100.0%	100.0%
(*n*)	(1,157)	(1,452)

Source: General Social Survey National Opinion Research Center 2016.

Evaluating Association

You will find when you read research reports and journal articles that social scientists usually make decisions about the existence and strength of association on the basis of more statistics than just a cross-tabulation table.

A measure of association is a type of descriptive statistic used to summarize the strength of an association. There are many measures of association, some of which are appropriate for variables measured at particular levels. One popular measure of association in cross-tabular analyses with variables measured at the ordinal level is gamma. As with many measures of association, the possible values of gamma vary from -1, meaning the variables are perfectly associated in an inverse direction; to 0, meaning there is no association of the type that gamma measures; to +1, meaning there is a perfect positive association of the type that gamma measures.

Exhibit 9.29 provides a rough guide to interpreting the value of a measure of association like gamma that can range from 0 to -1 and +1. For example, if the value of gamma is -.23, we could say that there is a *weak negative relationship* between the two variables. If the value of gamma is +.61, we could say that there is a *strong positive relationship* between the two variables. A value of 0 always means that there is no relationship (although this really means there is no relationship that this particular statistic can identify). This "rough guide" to interpretation must be modified for some particular measures of association, and your interpretations must consider the results of previous research and the particular methods you used to collect your data. For now, however, this rough guide will get you further along in your statistical interpretations.

Inferential statistics are used in deciding whether it is likely that an association exists in the larger population from which the sample was drawn. Even when the association between two variables is consistent with the researcher's hypothesis, it is possible that the association was just caused by the vagaries of sampling on a random basis (of course, the problem is even worse if the sample is not random). It is conventional in statistics to avoid concluding that an association exists in the population from which the sample was drawn unless the probability that the association was due to chance is less than 5%. In other words, a statistician typically will not conclude that an association exists between two variables unless he or she can be at least 95% confident that the association was not due to chance. This is the same type of logic that you learned about earlier in this chapter, which introduced the concept of 95% confidence limits for the mean. Estimation of the probability that an association is not due to chance will be based on one of several inferential statistics, chi-square being the one used in most cross-tabular analyses. The probability is customarily reported in a summary form such as $p < .05$, which can be translated as "The probability that the association was due to chance is less than 5 out of

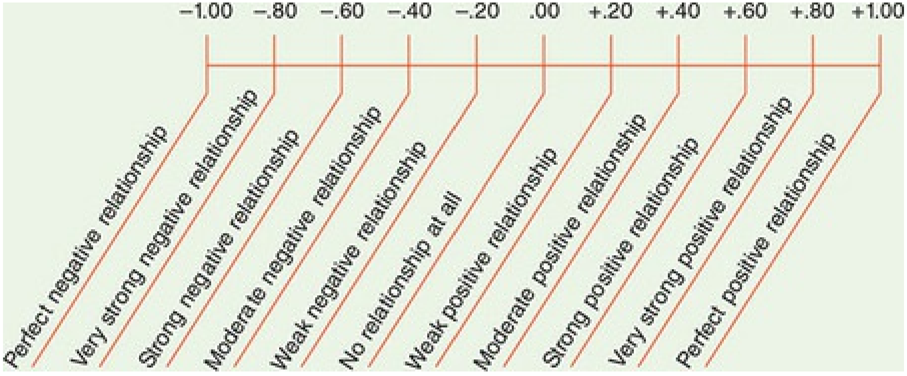
100 (5%).”

Measure of association: A type of descriptive statistic that summarizes the strength of an association.

Gamma: A measure of association that is sometimes used in cross-tabular analysis.

Chi-square: An inferential statistic used to test hypotheses about relationships between two or more variables in a cross-tabulation.

Exhibit 9.29 A Guide to Interpreting Strong and Weak Relationships


−1.00
−.80
−.60
−.40
−.20
.00
+.20
+.40
+.60
+.80
+1.00
Perfect negative relationship
Very strong negative relationship
Strong negative relationship
Moderate negative relationship
Weak negative relationship
No relationship at all
Weak positive relationship
Moderate positive relationship
Strong positive relationship
Very strong positive relationship
Perfect positive relationship

Source: Frankfort-Nachmias and Leon-Guerrero (2006:230). Reprinted with permission from SAGE Publications, Inc.

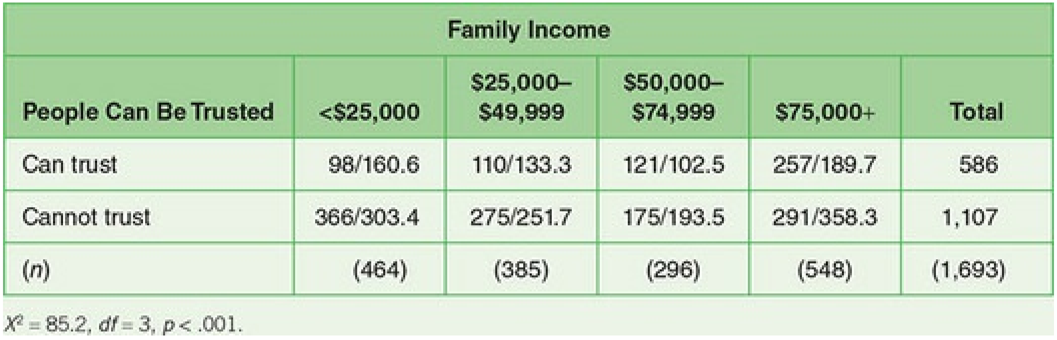
The tables in Exhibit 9.30 and 9.31 will help you understand the meaning of chi-square, without getting into the details of how it is calculated. Let’s propose, as our “null hypothesis,” that trust in other people has no association with family income. In that case, trust in people would be the same percentage for all columns in the table—the same as for the overall sample across all four income categories (34.6%). So if there were no association, we would expect on the basis of chance that 34.6% of the 464 people with family income below $25,000 (the first column) will say they can trust other people. Because 34.6% of 464 equals 160.6, that is the number of people we would “expect” to be trusting and of low income if only chance factors were at work. This is the *expected count* and it differs from the *actual count* of 98, leaving a *residual* of -62.6. This process is repeated with respect to each cell of the table. The larger the deviations of the expected from the observed counts in the various table cells, the less likely it is that the association is due only to chance. Chi-square is calculated with a formula that combines the residuals in each cell. SPSS then compares the value of chi-square to a table that indicates how likely it is in a table of the given size that this value could have been obtained on the basis of chance, given the “degrees of freedom” (*df*) in the table [(the number of rows -1) × (the number of columns -1)]. In the crosstab of family income and trust, the value of chi-square was 85.2 and the probability that a chi-square value of this magnitude was obtained on the basis of chance was less than 1 in 1,000 ($p < .001$). We could therefore feel confident that an association between these two variables exists in the U.S. adult population as a whole.

When the analyst feels reasonably confident (at least 95% confident) that an association was not due to chance, it is said that the association is statistically significant. Statistical significance means that an association is not likely to result from chance, according to some criterion set by the analyst. Convention (and the desire to avoid concluding that an association exists in the population when it doesn't) dictates that the criterion be a probability less than 5%.

Statistical significance: The mathematical likelihood that an association is due to chance, judged by a criterion set by the analyst (often that the probability is less than 5 out of 100 or $p < .05$).

But statistical significance is not everything. You may remember from Chapter 5 that sampling error decreases as sample size increases. For this same reason, an association is less likely to appear on the basis of chance in a larger sample than in a smaller sample. In a table with more than 1,000 cases, such as those involving the full 2016 GSS sample, the odds of a chance association are often very low indeed. For example, with our table based on 2,387 cases, the probability that the association between income and voting (see Exhibit 9.23) was due to chance was less than 1 in 1,000 ($p < .001$)! Nonetheless, the association in that table was weak, as indicated by a gamma of -.32. Even weak associations can be statistically significant with such a large random sample, which means that the analyst must be careful not to assume that just because a statistically significant association exists, it is therefore important. In a large sample, an association may be statistically significant but still be too weak to be substantively significant. All this boils down to another reason for evaluating carefully both the existence and the strength of an association.

Exhibit 9.30 Determining the Value of Chi-Square (actual/expected counts)

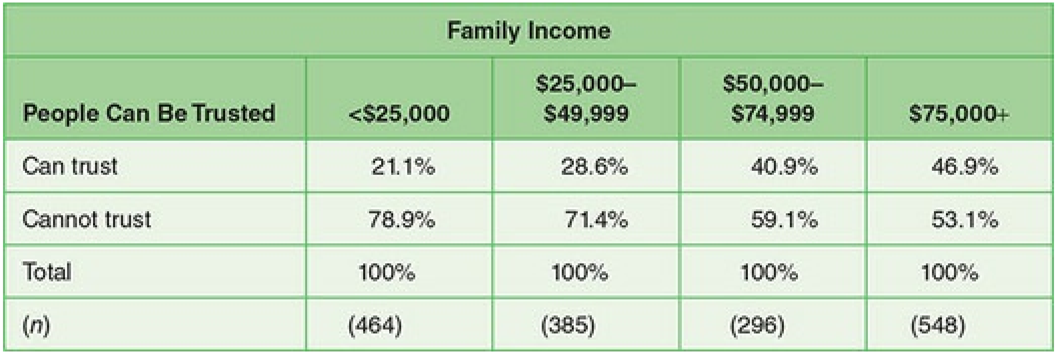


	Family Income				
People Can Be Trusted	**<$25,000**	**$25,000–$49,999**	**$50,000–$74,999**	**$75,000+**	**Total**
Can trust	98/160.6	110/133.3	121/102.5	257/189.7	586
Cannot trust	366/303.4	275/251.7	175/193.5	291/358.3	1,107
(*n*)	(464)	(385)	(296)	(548)	(1,693)

$X^2 = 85.2$, $df = 3$, $p < .001$.

Source: Based on output from General Social Survey National Opinion Research Center 2016.

Exhibit 9.31 Cross-Tabulation of Interpersonal Trust by Income



	Family Income			
People Can Be Trusted	**<$25,000**	**$25,000–$49,999**	**$50,000–$74,999**	**$75,000+**
Can trust	21.1%	28.6%	40.9%	46.9%
Cannot trust	78.9%	71.4%	59.1%	53.1%
Total	100%	100%	100%	100%
(*n*)	(464)	(385)	(296)	(548)

Source: General Social Survey National Opinion Research Center 2016.

Controlling for a Third Variable

Cross-tabulation can also be used to study the relationship between two variables while controlling for other variables. We will focus our attention on controlling for a third variable in this section, but I will say a bit about controlling for more variables at the section's end. We will examine three different uses for three-variable cross-tabulation: (1) identifying an intervening variable, (2) testing a relationship for spuriousness, and (3) specifying the conditions for a relationship. Each type of three-variable crosstab helps strengthen our understanding of the "focal relationship" involving our dependent and independent variables (Aneshensel 2002). Testing a relationship for possible spuriousness helps meet the nonspuriousness criterion for causality; this was the main focus of the Campbell and Horowitz (2016) test of the spuriousness of the apparent effect of college on sociopolitical attitudes (see "Research That Matters"). Identifying an intervening variable can help chart the causal mechanism by which variation in the independent variable influences variation in the dependent variable; Campbell and Horowitz (2016:41–42) considered three possible intervening variables as causal mechanisms connecting college attendance to sociopolitical attitudes, although they did not actually test them. Specifying the conditions when a relationship occurs can help improve our understanding of the nature of that relationship; a goal suggested by Campbell and Horowitz (2016:55–56) for future research that can take into account differences in the effects of college by gender, socioeconomic status, and historical era.

All three uses for three-variable cross-tabulation are aspects of elaboration analysis: the process of introducing control variables into a bivariate relationship to better understand the relationship (Davis 1985; Rosenberg 1968). We will examine the gamma and chi-square statistics for each table in this analysis.

> Elaboration analysis: The process of introducing a third variable into an analysis to better understand—to elaborate—the bivariate (two-variable) relationship under consideration. Additional control variables also can be introduced.

Intervening Variables

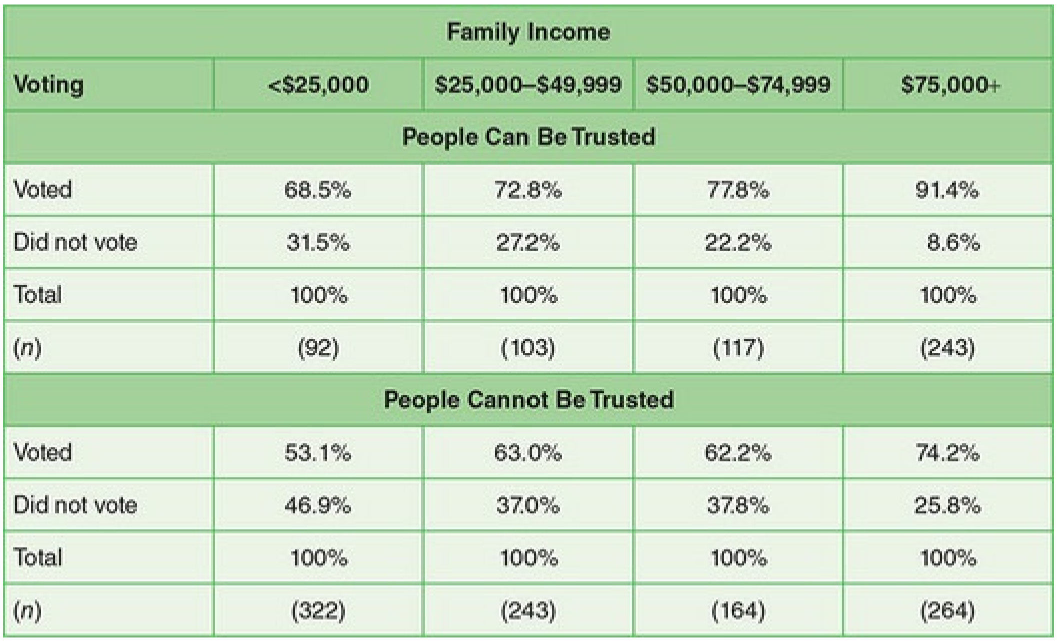
We will first complete our test of one of the implications of the causal model of voting in Exhibit 9.1: that trust (or efficacy) intervenes in the relationship between social status and voting. You already have seen that both income (one of our social status indicators) and trust in people are associated with the likelihood of voting. Both relationships are predicted by the model: so far, so good. You can also see in Exhibit 9.31 that trust is related to income: Higher income is associated positively with the belief that people can be trusted (gamma = -.35; $p < .001$). Another prediction of the model is confirmed. But to determine whether the trust variable is an intervening variable in this relationship, we must determine whether it explains (transmits) the influence of income on trust. We therefore examine the relationship between income and voting while controlling for the respondent's belief that people can be trusted.

According to the causal model, income (social status) influences voting (political participation) by influencing trust in people (our substitute for efficacy), which, in turn, influences voting. We can evaluate this possibility by reading the two subtables in Exhibit 9.32. Subtables such as those in Exhibit 9.32 describe the relationship between two variables within the discrete categories of one or more other control variables. The control variable in Exhibit 9.32 is trust in people, and the first subtable is the income-voting crosstab for only those respondents who believe that people can be trusted. The second subtable is for those respondents who believe that people can't be trusted. They are called subtables because together they make up the table in Exhibit 9.32. If trust in ordinary people intervened in the income–voting relationship, then the effect of controlling for this third variable would be to eliminate, or at least substantially reduce, this relationship—the distribution of voting would be the same for every income category in both subtables in Exhibit 9.32.

Subtables: Tables describing the relationship between two variables within the discrete categories of one or more other control variables.

A quick inspection of the subtables in Exhibit 9.32 reveals that trust in people does not intervene in the relationship between income and voting. There is only a modest difference in the strength of the income–voting association in the subtables (as reflected in the value of gamma, which is -.435 in the first subtable and -.249 in the second). In both subtables, the likelihood that respondents voted rose with their incomes. Of course, this finding does not necessarily mean that the causal model was wrong. This one measure is a measure of trust in people, which is not the same as the widely studied concept of political efficacy; a better measure, from a different survey, might function as an intervening variable. But for now, we should be less confident in the model.

Exhibit 9.32 Voting in 2012 by Family Income by Interpersonal Trust



Voting	Family Income <$25,000	 $25,000–$49,999	 $50,000–$74,999	 $75,000+
People Can Be Trusted				
Voted	68.5%	72.8%	77.8%	91.4%
Did not vote	31.5%	27.2%	22.2%	8.6%
Total	100%	100%	100%	100%
(*n*)	(92)	(103)	(117)	(243)
People Cannot Be Trusted				
Voted	53.1%	63.0%	62.2%	74.2%
Did not vote	46.9%	37.0%	37.8%	25.8%
Total	100%	100%	100%	100%
(*n*)	(322)	(243)	(164)	(264)

Source: General Social Survey National Opinion Research Center 2016.

Extraneous Variables

Another reason for introducing a third variable into a bivariate relationship is to see whether that relationship is spurious because of the influence of an extraneous variable (see Chapter 6)—a variable that influences both the independent and dependent variables, creating an association between them that disappears when the extraneous variable is controlled. Ruling out possible extraneous variables will help strengthen considerably the conclusion that the relationship between the independent and dependent variables is causal, particularly if all the variables that seem to have the potential for creating a spurious relationship can be controlled.

One variable that might create a spurious relationship between income and voting is education. You have already seen that the likelihood of voting increases with income. Is it not possible, though, that this association is spurious because of the effect of education? Education, after all, is associated with both income and voting, and we might surmise that it is what students learn in school about civic responsibility that increases voting, not income itself. Exhibit 9.33 diagrams this possibility, and Exhibit 9.34 shows the bivariate associations between education and voting, and education and income. As the model in Exhibit 9.33 predicts, education is associated with both income and voting. So far, so good. If education actually does create a spurious relationship between income and voting, there should be no association between income and voting after controlling for education. Because we are using crosstabs, this means there should be no association in any of the income–voting subtables for any value of education.

The trivariate cross-tabulation in Exhibit 9.35 shows that the relationship between voting and income is not spurious because of the effect of education; if it were, an association between voting and family income wouldn't appear in any of the subtables—somewhat like the first two subtables, in which gamma is only .01 and .08, respectively.

The association between family income and voting is higher in the other two subtables in Exhibit 9.35, for respondents with some college or a college education. The strength of that association as measured by gamma is -.199 for those with some college, and it is -.342 among college graduates. So our hypothesis—that income as a social status indicator leads to higher rates of voting—does not appear to be spurious because of the effect of education. The next section elaborates on the more complex pattern that we found.

Exhibit 9.33 A Causal Model of a Spurious Effect

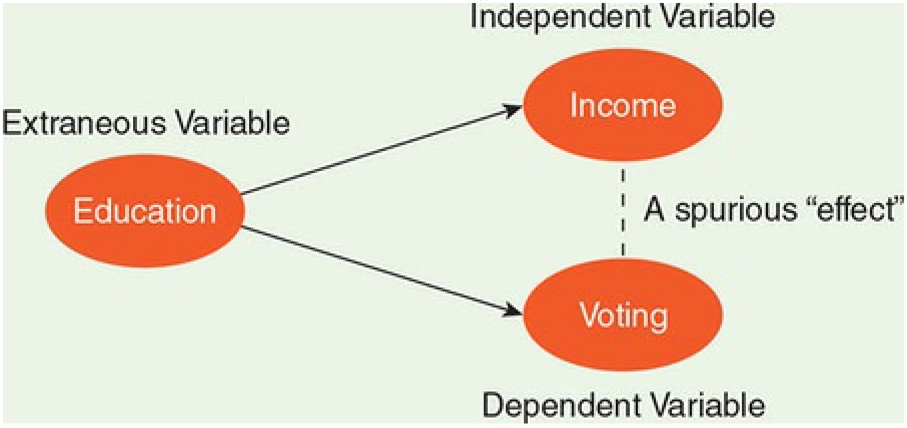
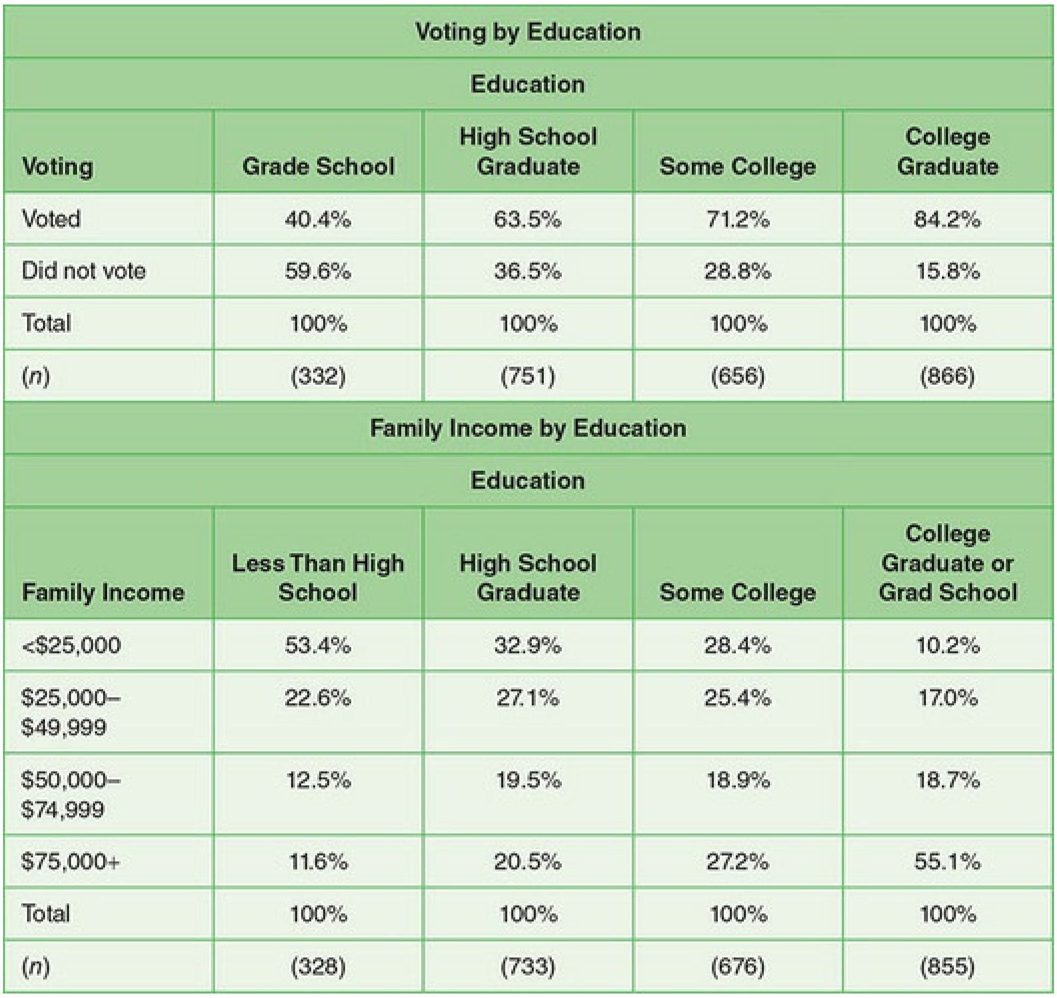

Independent Variable
Extraneous Variable
Income
Education
A spurious "effect"
Voting
Dependent Variable

Exhibit 9.34 Voting in 2012 by Education and Income by Education



Voting by Education

	Education			
Voting	Grade School	High School Graduate	Some College	College Graduate
Voted	40.4%	63.5%	71.2%	84.2%
Did not vote	59.6%	36.5%	28.8%	15.8%
Total	100%	100%	100%	100%
(*n*)	(332)	(751)	(656)	(866)

Family Income by Education

	Education			
Family Income	Less Than High School	High School Graduate	Some College	College Graduate or Grad School
<$25,000	53.4%	32.9%	28.4%	10.2%
$25,000–$49,999	22.6%	27.1%	25.4%	17.0%
$50,000–$74,999	12.5%	19.5%	18.9%	18.7%
$75,000+	11.6%	20.5%	27.2%	55.1%
Total	100%	100%	100%	100%
(*n*)	(328)	(733)	(676)	(855)

Source: General Social Survey National Opinion Research Center 2016.

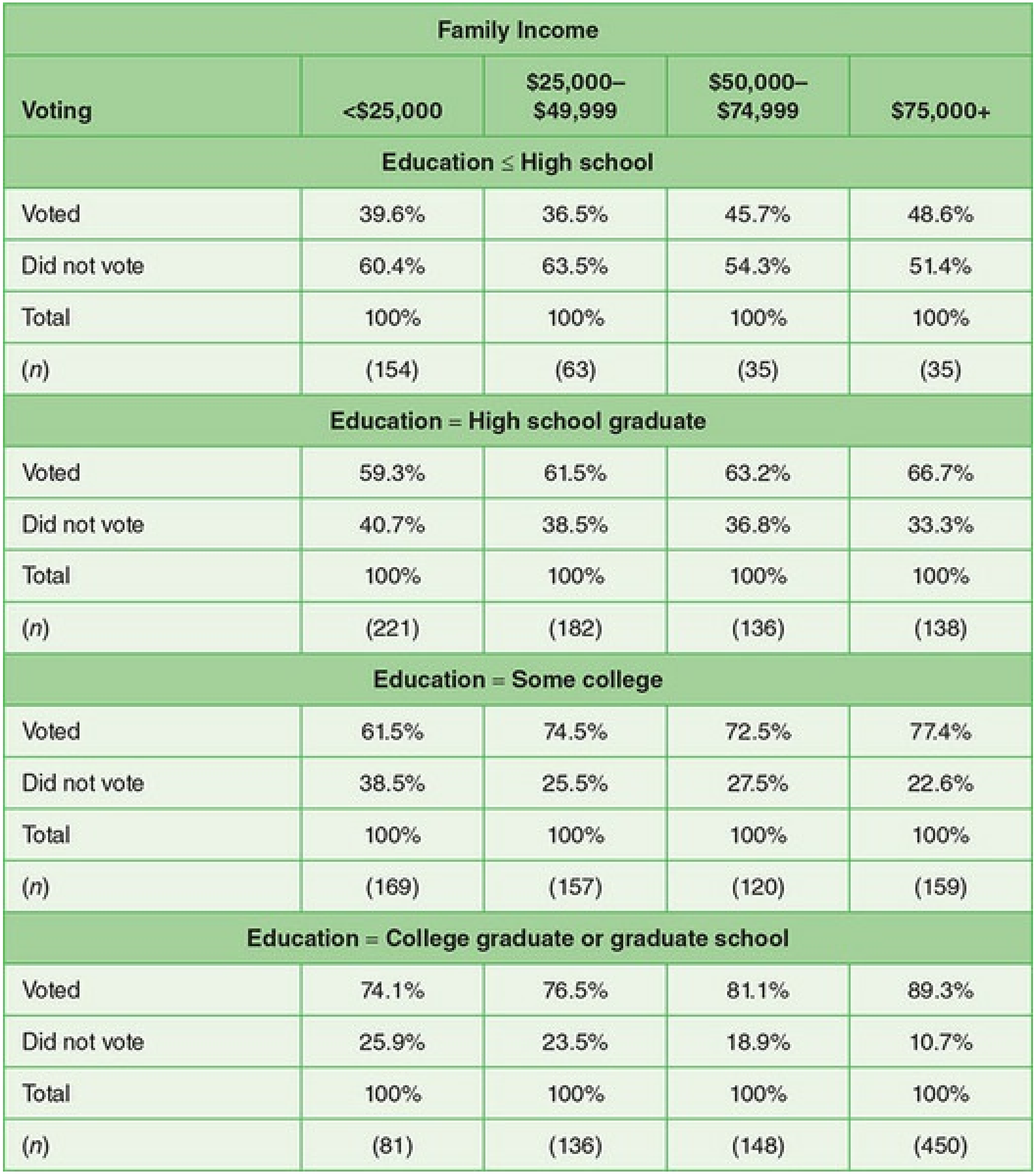
Specification

By adding a third variable to an evaluation of a bivariate relationship, the data analyst can also specify the conditions under which the bivariate relationship occurs. A specification occurs when the association between the independent and dependent variables varies across the categories of one or more other control variables. This is what we just found in Exhibit 9.35: There is almost no association between income and voting for those with a high school education or less, but there is a moderate association for the higher educational categories.

Specification: A type of relationship involving three or more variables in which the association between the independent and dependent variables varies across the categories of one or more other control variables.

The subtables in Exhibit 9.36 allow an evaluation of whether race specifies the effect of income on voting, as suggested by previous research. The percentages who voted in each of the family income categories vary less among African Americans (gamma = -.20) and respondents who identify themselves as members of other minority groups (gamma = -.30) than among whites (gamma = .37). Race, therefore, does appear to specify the association between income and voting: The likelihood of African American and other minority respondents having voted varies much less with their family income than it does among whites. The lower rate of voting among members of other minority groups is itself of interest; investigation of the reason for this would make for an interesting contribution to the literature. Is it because members of other minority groups are more likely to be recent immigrants and so less engaged in the U.S. political system than whites or African Americans? Can you think of other possibilities?

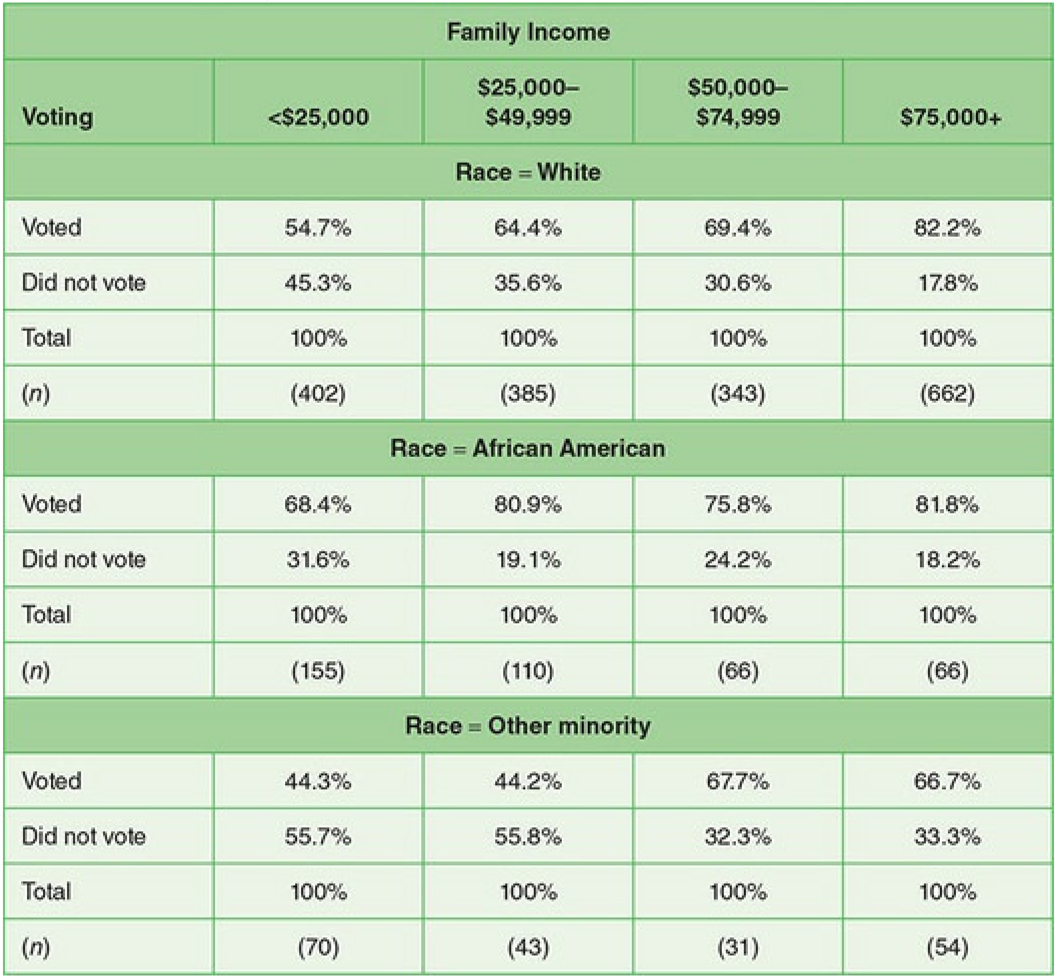
Exhibit 9.35 Voting in 2012 by Income and Education



	Family Income			
Voting	**<$25,000**	**$25,000–$49,999**	**$50,000–$74,999**	**$75,000+**
Education ≤ High school				
Voted	39.6%	36.5%	45.7%	48.6%
Did not vote	60.4%	63.5%	54.3%	51.4%
Total	100%	100%	100%	100%
(*n*)	(154)	(63)	(35)	(35)
Education = High school graduate				
Voted	59.3%	61.5%	63.2%	66.7%
Did not vote	40.7%	38.5%	36.8%	33.3%
Total	100%	100%	100%	100%
(*n*)	(221)	(182)	(136)	(138)
Education = Some college				
Voted	61.5%	74.5%	72.5%	77.4%
Did not vote	38.5%	25.5%	27.5%	22.6%
Total	100%	100%	100%	100%
(*n*)	(169)	(157)	(120)	(159)
Education = College graduate or graduate school				
Voted	74.1%	76.5%	81.1%	89.3%
Did not vote	25.9%	23.5%	18.9%	10.7%
Total	100%	100%	100%	100%
(*n*)	(81)	(136)	(148)	(450)

Source: General Social Survey National Opinion Research Center 2016.

Exhibit 9.36 Voting in 2012 by Income and Race



Family Income				
Voting	**<$25,000**	**$25,000–$49,999**	**$50,000–$74,999**	**$75,000+**
Race = White				
Voted	54.7%	64.4%	69.4%	82.2%
Did not vote	45.3%	35.6%	30.6%	17.8%
Total	100%	100%	100%	100%
(*n*)	(402)	(385)	(343)	(662)
Race = African American				
Voted	68.4%	80.9%	75.8%	81.8%
Did not vote	31.6%	19.1%	24.2%	18.2%
Total	100%	100%	100%	100%
(*n*)	(155)	(110)	(66)	(66)
Race = Other minority				
Voted	44.3%	44.2%	67.7%	66.7%
Did not vote	55.7%	55.8%	32.3%	33.3%
Total	100%	100%	100%	100%
(*n*)	(70)	(43)	(31)	(54)

Source: General Social Survey National Opinion Research Center 2016.

I should add one important caution about constructing tables involving three or more variables. Because the total number of cells in the subtables becomes large as the number of categories of the control (third) variable increases, the number of cases that each cell percentage is based on will become correspondingly small. This effect has two important consequences. First, the number of comparisons that must be made to identify the patterns in the table as a whole becomes substantial—and the patterns may become too complex to make much sense of them. Second, as the number of cases per category decreases, the odds that the distributions within each category vary because of chance become greater. This problem of having too many cells and too few cases can be lessened by making sure that the control variable has only a few categories and by drawing a large sample, but often neither of these steps will be sufficient to resolve the problem completely.

Careers and Research

Claire Wulf Winiarek, MA, Director of Collaborative Policy Engagement

Claire Wulf Winiarek didn't set her sights on research methods as an undergraduate in political science and international relations at Mary Baldwin University, nor as a masters student at Old Dominion University; her goal was to make a difference in public affairs. It still is. She is currently vice president of public policy at Magellan Health, one of *Fortune* magazine's World's Most Admired Companies based in Scottsdale, Arizona. Her previous positions include managing director of Public Policy for Anthem, Inc., a *Fortune* 50 health insurer; director of public policy and research at Amerigroup Corporation; staffer to a Virginia member of the U.S. House of Representatives, and coordinating grassroots human rights advocacy for Amnesty International's North Africa Regional Action Network.

Early in her career, Winiarek was surprised by the frequency with which she found herself leveraging research methods. Whether she is analyzing draft legislation and proposed regulations, determining next year's department budget, or estimating potential growth while making the case for a new program, Winiarek has found that a strong foundation in research methods shapes her success. The increasing reliance of government and its private-sector partners on data and evidence-based decision making continues to increase the importance of methodological expertise.

Policy work informed by research has made for a very rewarding career:

> The potential for meaningful impact in the lives of every day Americans is very real at the nexus of government and the private sector. Public policy, and how policy works in practice, has significant societal impact. I feel fortunate to help advance that nexus in a way that is informed not only by practice, evidence, and research, but also by the voice of those impacted.

Winiarek's advice for students seeking a career like hers is clear:

> The information revolution is impacting all industries and sectors, as well as government and our communities. With this ever growing and ever richer set of information, today's professionals must have the know-how to understand and apply this data in a meaningful way. Research methods will create the critical and analytical foundation to meet the challenge, but internships or special research projects in your career field will inform that foundation with practical experience. Always look for that connection between research and reality.

Regression Analysis

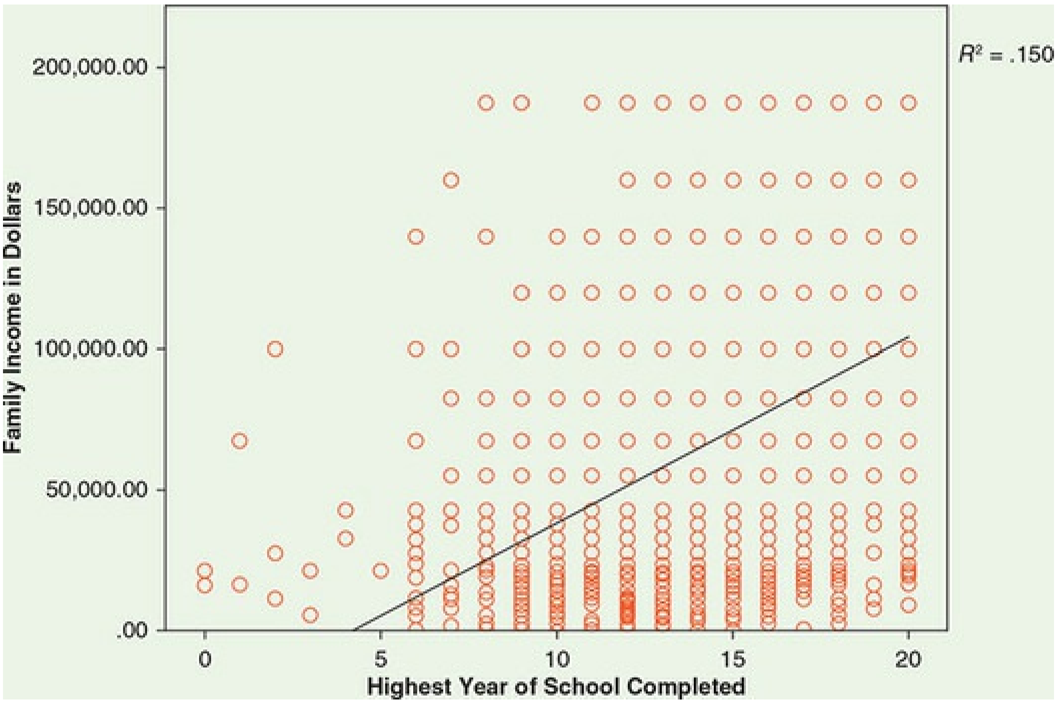
My goal in introducing you to cross-tabulation has been to help you think about the association between variables and to give you a relatively easy tool for describing association. To read most statistical reports and to conduct more sophisticated analyses of social data, you will have to extend your statistical knowledge. Many statistical reports and articles published in social science journals use a statistical technique called regression analysis or correlational analysis to describe the association between two or more quantitative variables. The terms actually refer to different aspects of the same technique. Statistics based on regression and correlation are used frequently in social science and have many advantages over cross-tabulation—as well as some disadvantages.

Regression analysis: A statistical technique for characterizing the pattern of a relationship between two quantitative variables in terms of a linear equation and for summarizing the strength of this relationship in terms of its deviation from that linear pattern.

Correlational analysis: A statistical technique that summarizes the strength of a relationship between two quantitative variables in terms of its adherence to a linear pattern.

I give you only an overview of this approach here. Take a look at Exhibit 9.37. It's a plot, termed a *scatterplot,* of the relationship in the GSS 2016 sample between years of education and family income. You can see that I didn't collapse the values of either of these variables into categories, as I had to do to use them in the preceding cross-tabular analysis. Instead, the scatterplot shows the location of each case in the data relative to years of education (the horizontal axis) and income level (the vertical axis).

Exhibit 9.37 Family Income by Highest Year of School Completed



Source: General Social Survey National Opinion Research Center 2016.

You can see that the data points in the scatterplot tend to run from the lower left to the upper right of the chart, indicating a positive relationship: The more the years of education, the higher the family income. The line drawn through the points is the regression line. The regression line summarizes this positive relationship between years of education, which is the independent variable (often simply termed *X* in regression analysis), and family income, the dependent variable (often simply termed *Y* in regression analysis). This regression line is the "best fitting" straight line for this relationship—it is the line that lies closest to all the points in the chart, according to certain criteria. But you can easily see that quite a few points are pretty far from the regression line.

How well does the regression line fit the points? In other words, how close does the regression line come to the points? (Actually, it's the square of the vertical distance, on the *y*-axis, between the points and the regression line that is used as the criterion.) The correlation coefficient, also called *Pearson's r,* or just *r,* gives one answer to that question. The value of *r* for this relationship is .39, which indicates a moderately strong positive linear relationship (if it were a negative relationship, *r* would have a negative sign). The value of *r* is 0 when there is absolutely no linear relationship between the two variables, and it is 1 when all the points representing all the cases lie exactly on the regression line (which would mean that the regression line describes the relationship perfectly).

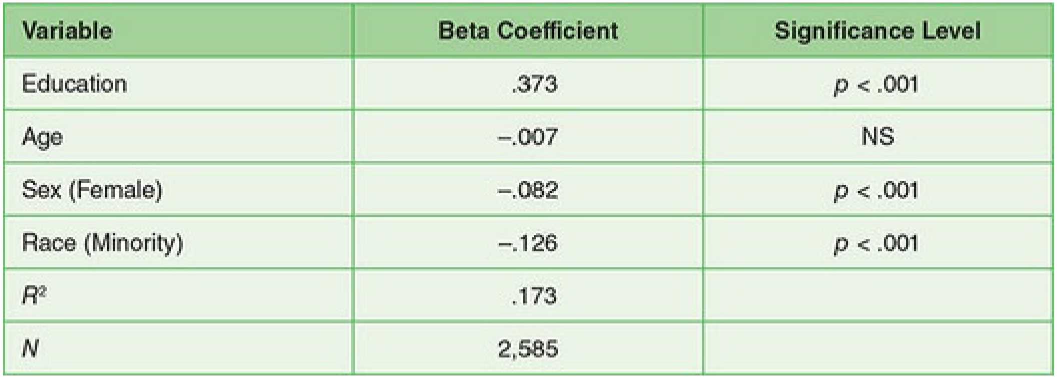
So the correlation coefficient does for a scatterplot such as this what gamma does for a cross-tabulation table: It is a summary statistic that tells us about the strength of the association between the two variables. Values of *r* close to 0 indicate that the relationship is weak; values of *r* close to ±1 indicate the relationship is strong—in between there is a lot of room for judgment. You will learn in a statistics course that r^2 is often used instead of *r*.

Correlation coefficient: A summary statistic that varies from 0 to 1 or –1, with 0 indicating the absence of a linear relationship between two quantitative variables and 1 or –1 indicating that the relationship is completely described by the line representing the regression of the dependent variable on the independent variable.

You can also use correlation coefficients and regression analysis to study simultaneously the association between three or more variables. In such a *multiple regression analysis,* you could test to see whether several other variables in addition to education are associated simultaneously with family income—that is, whether the variables have independent effects on family income. As an example, Exhibit 9.38 presents the key statistics obtained in a multiple regression analysis I conducted with the GSS 2016 data to do just that: I regressed family income on years of schooling, age, sex, and race (dichotomized).

First, look at the numbers under the "Beta Coefficient" heading. Beta coefficients are standardized statistics that indicate how strong the linear association is between the dependent variable (family income, in this case) and each independent variable, while the other independent variables are controlled. Like the correlation coefficient (*r*), values of beta range from 0, when there is no linear association, to ±1.0, when the association falls exactly on a straight line. You can see in the beta column that education has a moderate positive independent association with family income, whereas sex and race have a weak association and age has no (linear) association with family income. In the "Significance Level" column, you can see that each of the three effects is statistically significant at the .001 level. You learn from the summary statistic, R^2 (*r*-squared), that the four independent variables together explain, or account for, 17% of the total variation in family income.

Exhibit 9.38 Multiple Regression of Determinants of Family Income



Variable	Beta Coefficient	Significance Level
Education	.373	$p < .001$
Age	–.007	NS
Sex (Female)	–.082	$p < .001$
Race (Minority)	–.126	$p < .001$
R^2	.173	
N	2,585	

Source: General Social Survey National Opinion Research Center 2016.

You will need to learn more about when correlation coefficients and regression analysis are appropriate (e.g., both variables have to be quantitative, and the relationship has to be linear [not curvilinear]). But that's for another time and place. To learn more about correlation coefficients and regression analysis, you should take an entire statistics course. For now, this short introduction will enable you to make sense of more of the statistical analyses you find in research articles. You can also learn more about these techniques with the tutorials on the text's study site.

Performing Meta-Analyses

Meta-analysis is a quantitative method for identifying patterns in findings across multiple studies of the same research question (Cooper and Hedges 1994). Unlike a traditional literature review, which describes previous research studies verbally, meta-analyses treat previous studies as cases, whose features are measured as variables and then analyzed statistically. It is like conducting a survey in which the "respondents" are previous studies. Meta-analysis shows how evidence about social processes varies across research studies. If the methods used in these studies varied, then meta-analysis can describe how this variation affected the study findings. If social contexts varied across the studies, then meta-analysis will indicate how social context affected the study findings. You have already learned in this chapter about most of the statistics that are used in meta-analyses.

Meta-analysis: The quantitative analysis of findings from multiple studies.

Meta-analysis can be used when a number of studies have attempted to answer the same research question with similar quantitative methods, most often experiments. Meta-analysis is not appropriate for evaluating results from qualitative studies or from multiple studies that used different methods or measured different dependent variables. It is also not very sensible to use meta-analysis to combine study results when the original case data from these studies are available and can actually be combined and analyzed together (Lipsey and Wilson 2001). Meta-analysis is a technique for combination and statistical analysis of published research reports.

After a research problem is formulated based on the findings of prior research, the literature must be searched systematically to identify the entire population of relevant studies (see Chapter 2). Typically, multiple bibliographic databases are used; some researchers also search for relevant dissertations and conference papers. Once the studies are identified, their findings, methods, and other features are coded (e.g., sample size, location of sample, and strength of the association between the independent and dependent variables). Eligibility criteria must be specified carefully to determine which studies to include and which to omit as too different. Mark Lipsey and David Wilson (2001:16–21) suggested that eligibility criteria include the following:

- *Distinguishing features.* This includes the specific intervention tested and perhaps the groups compared.
- *Research respondents.* The pertinent characteristics of the research respondents (subject sample) who provided study data must be similar to those of the population about which generalization is sought.

- *Key variables.* These must be sufficient to allow tests of the hypotheses of concern and controls for likely additional influences.
- *Research methods.* Apples and oranges cannot be directly compared, but some trade-off must be made between including the range of studies about a research question and excluding those that are so different in their methods as not to yield comparable data.
- *Cultural and linguistic range.* If the study population is going to be limited to English-language publications, or limited in some other way, this must be acknowledged, and the size of the population of relevant studies in other languages should be estimated.
- *Time frame.* Social processes relevant to the research question may have changed for reasons such as historical events or the advent of new technologies, so temporal boundaries around the study population must be considered.
- *Publication type.* It must be determined whether the analysis will focus only on published reports in professional journals, or include dissertations and unpublished reports.

Statistics are then calculated to identify the average effect of the independent variable on the dependent variable, as well as the effect of methodological and other features of the studies (Cooper and Hedges 1994). The effect size statistic is the key to capturing the association between the independent and dependent variables across multiple studies. The effect size statistic is a standardized measure of association—often the difference between the mean of the experimental group and the mean of the control group on the dependent variable, adjusted for the average variability in the two groups (Lipsey and Wilson 2001).

Effect size: A standardized measure of association—often the difference between the mean of the experimental group and the mean of the control group on the dependent variable, adjusted for the average variability in the two groups.

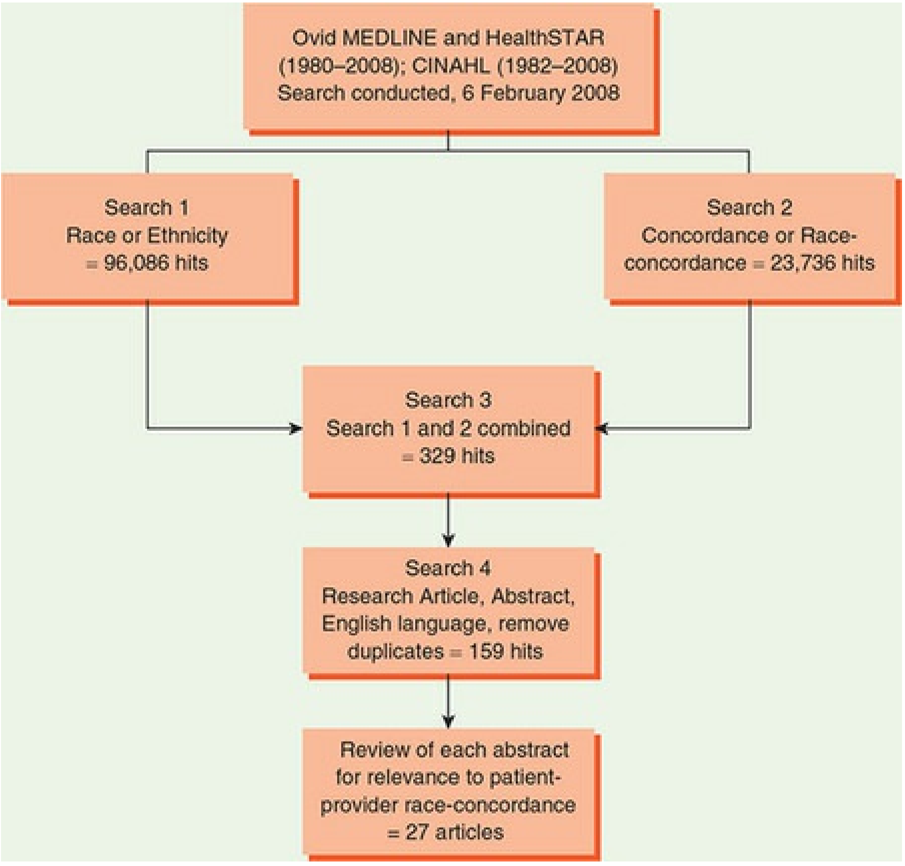
The meta-analytic approach to synthesizing research findings can result in much more generalizable findings than those obtained with just one study. Methodological weaknesses in the studies included in the meta-analysis are still a problem, however; only when other studies without particular methodological weaknesses are included can we estimate effects with some confidence. In addition, before we can place any confidence in the results of a meta-analysis, we must be confident that all (or almost all) relevant studies were included and that the information we need to analyze was included in all (or most) of the studies (Matt and Cook 1994).

Case Study: Patient–Provider Race Concordance and Minority Health Outcomes

Do minority patients have better health outcomes when they receive treatment from a provider of the same race or ethnicity? Salimah Meghani and other researchers in nursing at the University of Pennsylvania and other Pennsylvania institutions sought to answer this question with a meta-analysis of published research. Their research report illustrates the key steps in a meta-analysis (Meghani et al. 2009).

They began their analysis with a comprehensive review of published research that could be located in three health-related bibliographic databases with searches for English-language research articles linked to the key words *race, ethnicity, concordance,* or *race concordance.* This search identified 159 articles; after reading the abstracts of these articles, 27 were identified that had investigated a research question about the effect of patient–provider race concordance on minority patients' health outcomes (see Exhibit 9.39).

Exhibit 9.39 Identification and Selection of Articles in Meta-Analysis

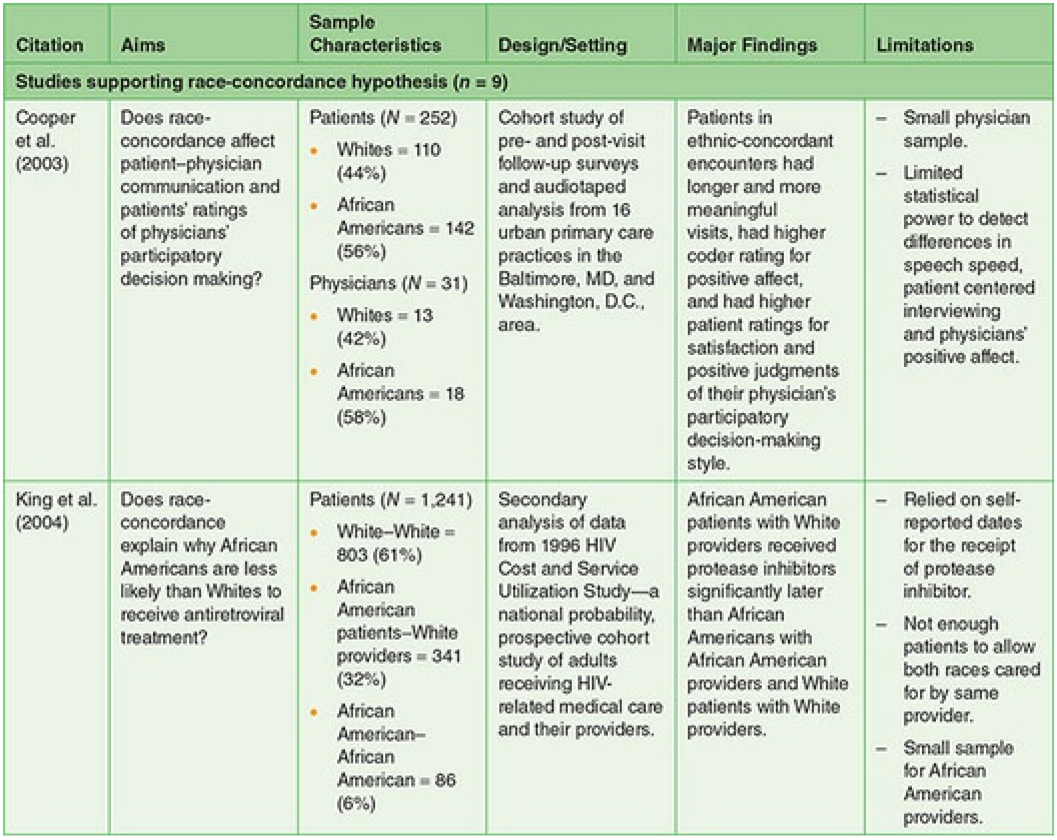

Ovid MEDLINE and HealthSTAR (1980–2008); CINAHL (1982–2008) Search conducted, 6 February 2008
Search 1 Race or Ethnicity = 96,086 hits
Search 2 Concordance or Race-concordance = 23,736 hits
Search 3 Search 1 and 2 combined = 329 hits
Search 4 Research Article, Abstract, English language, remove duplicates = 159 hits
Review of each abstract for relevance to patient-provider race-concordance = 27 articles

Source: Meghani et al. (2009:109). Ethnicity and Health.

Meghani and her coauthors then summarized the characteristics and major findings of the selected studies (see Exhibit 9.40). Finally, each study was classified according to the health outcome(s) examined and its findings about the effect of race concordance on each outcome (see Exhibit 9.41). Because only 9 of the 27 studies provided support for a positive effect of race concordance on outcomes, and in many of these studies the effects were modest, Meghani and her coauthors concluded that patient–provider racial concordance had little relevance to health care outcomes.

Can you see why the systematic literature reviews you learned about in Chapter 2 often include a statistical meta-analysis of the findings from the studies reviewed? It is an excellent way to summarize the central findings in prior research and to highlight the extent to which these findings have been consistent. A meta-analysis also reminds us that the results of a single study have to be interpreted in relation to the larger body of research findings.

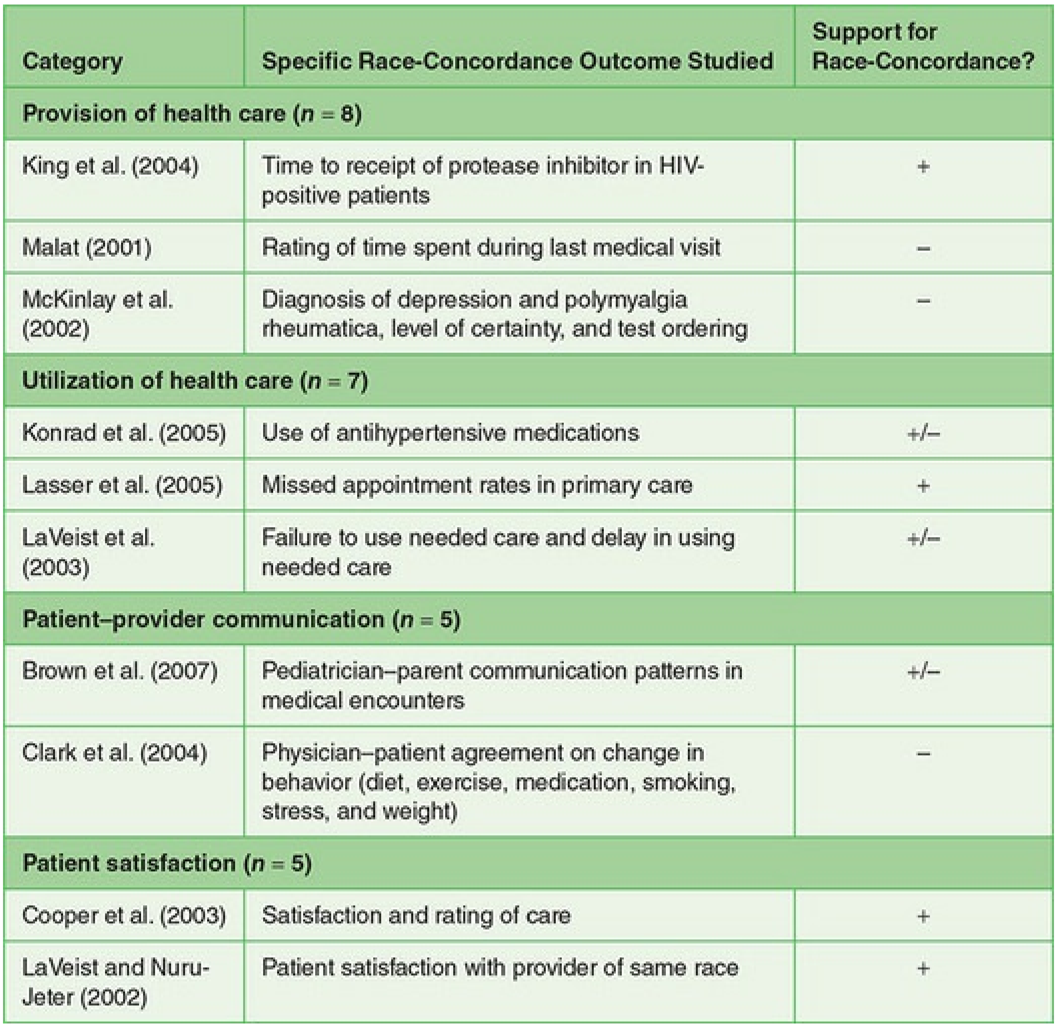
Exhibit 9.40 Summary of Studies in Meta-Analysis



Citation	Aims	Sample Characteristics	Design/Setting	Major Findings	Limitations
Studies supporting race-concordance hypothesis (*n* = 9)					
Cooper et al. (2003)	Does race-concordance affect patient–physician communication and patients' ratings of physicians' participatory decision making?	Patients (*N* = 252) • Whites = 110 (44%) • African Americans = 142 (56%) Physicians (*N* = 31) • Whites = 13 (42%) • African Americans = 18 (58%)	Cohort study of pre- and post-visit follow-up surveys and audiotaped analysis from 16 urban primary care practices in the Baltimore, MD, and Washington, D.C., area.	Patients in ethnic-concordant encounters had longer and more meaningful visits, had higher coder rating for positive affect, and had higher patient ratings for satisfaction and positive judgments of their physician's participatory decision-making style.	– Small physician sample. – Limited statistical power to detect differences in speech speed, patient centered interviewing and physicians' positive affect.
King et al. (2004)	Does race-concordance explain why African Americans are less likely than Whites to receive antiretroviral treatment?	Patients (*N* = 1,241) • White–White = 803 (61%) • African American patients–White providers = 341 (32%) • African American–African American = 86 (6%)	Secondary analysis of data from 1996 HIV Cost and Service Utilization Study—a national probability, prospective cohort study of adults receiving HIV-related medical care and their providers.	African American patients with White providers received protease inhibitors significantly later than African Americans with African American providers and White patients with White providers.	– Relied on self-reported dates for the receipt of protease inhibitor. – Not enough patients to allow both races cared for by same provider. – Small sample for African American providers.

Source: Meghani et al. (2009:111). Ethnicity and Health.

Exhibit 9.41 Classification of Outcomes in Meta-Analysis



Category	Specific Race-Concordance Outcome Studied	Support for Race-Concordance?
Provision of health care (*n* = 8)		
King et al. (2004)	Time to receipt of protease inhibitor in HIV-positive patients	+
Malat (2001)	Rating of time spent during last medical visit	–
McKinlay et al. (2002)	Diagnosis of depression and polymyalgia rheumatica, level of certainty, and test ordering	–
Utilization of health care (*n* = 7)		
Konrad et al. (2005)	Use of antihypertensive medications	+/–
Lasser et al. (2005)	Missed appointment rates in primary care	+
LaVeist et al. (2003)	Failure to use needed care and delay in using needed care	+/–
Patient–provider communication (*n* = 5)		
Brown et al. (2007)	Pediatrician–parent communication patterns in medical encounters	+/–
Clark et al. (2004)	Physician–patient agreement on change in behavior (diet, exercise, medication, smoking, stress, and weight)	–
Patient satisfaction (*n* = 5)		
Cooper et al. (2003)	Satisfaction and rating of care	+
LaVeist and Nuru-Jeter (2002)	Patient satisfaction with provider of same race	+

Source: Meghani et al. (2009:123). Ethnicity and Health.

Analyzing Data Ethically: How Not to Lie About Relationships

When the data analyst begins to examine relationships between variables in some real data, social science research becomes most exciting. The moment of truth, it would seem, has arrived. Either the hypotheses are supported or they are not. But this is actually a time to proceed with caution and to evaluate the analyses of others with even more caution. Once large data sets are entered into a computer, it becomes very easy to check out a great many relationships; when relationships are examined between three or more variables at a time, the possibilities become almost endless. In fact, regression analysis (extended to what is termed *multiple regression analysis*) allows a researcher to test easily for many relationships at the same time.

This range of possibilities presents a great hazard for data analysis. It becomes tempting to search around in the data until something interesting emerges. Rejected hypotheses are forgotten in favor of highlighting what's going on in the data. It's not wrong to examine data for unanticipated relationships; the problem is that inevitably some relationships between variables will appear just on the basis of chance association alone. If you search hard and long enough, it will be possible to come up with something that really means nothing.

A reasonable balance must be struck between deductive data analysis to test hypotheses and inductive analysis to explore patterns in a data set. Hypotheses formulated in advance of data collection must be tested as they were originally stated; any further analyses of these hypotheses that involve a more exploratory strategy must be labeled in research reports as such. Serendipitous findings do not need to be ignored, but they must be reported as such. Subsequent researchers can try to test deductively the ideas generated by our explorations.

We also have to be honest about the limitations of using survey data to test causal hypotheses. The usual practice for those who seek to test a causal hypothesis with nonexperimental survey data is to test for the relationship between the independent and dependent variables, controlling for other variables that might possibly create a spurious relationship. This is what we did by examining the relationship between income and voting while controlling for education (Exhibit 9.35). (These subtables show that education specifies the relationship between family income and voting—there is no relationship for those with only a grade school education, but the relationship exists for those who finished high school and those who attended college. Education does not explain the income–voting relationship.)

But finding that a hypothesized relationship is not altered by controlling for just one variable does not establish that the relationship is causal—nor does controlling for two, three, or many more variables. There always is a possibility that some other variable that we did not think to control, or that was not even measured in the survey, has produced a spurious relationship between the independent and dependent variables in our hypothesis (Lieberson 1985). We have to think about the possibilities and be cautious in our causal conclusions.

It is also important to understand the statistical techniques we are using and to use them appropriately. In particular, the analyst who uses regression analysis has to make a number of assumptions about the variables in the analysis; when these assumptions are violated, regression results can be very misleading. (You just might want to rush right out and buy this statistics text at this point: Frankfort-Nachmias, Chava and Anna Leon-Guerrero. 2015. *Social Statistics for a Diverse Society,* 7th ed. Thousand Oaks, CA: Sage.)

Conclusions

This chapter has demonstrated how a researcher can describe social phenomena, identify relationships between them, explore the reasons for these relationships, and test hypotheses about them. Statistics provide a remarkably useful tool for developing our understanding of the social world, a tool that we can use both to test our ideas and to generate new ones.

Unfortunately, to the uninitiated, the use of statistics can seem to end debate right there—you can't argue with the numbers. But you now know better than that. The numbers will be worthless if the methods used to generate the data are not valid; and the numbers will be misleading if they are not used appropriately, considering the type of data to which they are applied. And even assuming valid methods and proper use of statistics, there's one more critical step because the numbers do not speak for themselves. Ultimately, it is how we interpret and report the statistics that determines their usefulness.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Bar chart 318
Base number (*N*) 322
Bimodal 328
Central tendency 316
Chi-square 343
Coding 316
Combined frequency display 325
Compressed frequency display 326
Contingency table 337
Correlational analysis 351
Correlation coefficient 352
Cross-tabulation (crosstab) 337
Curvilinear 342
Data cleaning 316
Data entry 315
Descriptive statistics 313
Effect size 354
Elaboration analysis 345
Frequency distribution 322
Frequency polygon 318
Gamma 343
Histogram 318
Interquartile range 333
Marginal distribution 338
Mean 328
Measure of association 343
Median 328
Meta-analysis 353
Mode 327
Monotonic 342
Normal distribution 334
Outlier 333
Percentages 338
Precoding 316
Probability average 327
Quartiles 333
Range 332
Regression analysis 351
Skewness 316

Specification 348
Standard deviation 333
Statistical significance 344
Subtables 346
Unimodal 328
Variability 316
Variance 333

Highlights

- Data entry options include direct collection of data through a computer, use of scannable data entry forms, and use of data entry software. All data should be cleaned during the data entry process.
- Use of secondary data can save considerable time and resources but may limit data analysis possibilities.
- Bar charts, histograms, and frequency polygons are useful for describing the shape of distributions. Care must be taken with graphic displays to avoid distorting a distribution's apparent shape.
- Frequency distributions display variation in a form that can be easily inspected and described. Values should be grouped in frequency distributions in a way that does not alter the shape of the distribution. Following several guidelines can reduce the risk of problems.
- Some of the data in many reports can be displayed more efficiently by using combined and compressed statistical displays.
- Summary statistics often are used to describe the central tendency and variability of distributions. The appropriateness of the mode, mean, and median vary with a variable's level of measurement, the distribution's shape, and the purpose of the summary.
- The variance and standard deviation summarize variability around the mean. The interquartile range is usually preferable to the range to indicate the interval spanned by cases because of the effect of outliers on the range. The degree of skewness of a distribution is usually described in words rather than with a summary statistic.
- Honesty and openness are the key ethical principles that should guide data summaries.
- Cross-tabulations should normally be percentaged within the categories of the independent variable. A cross-tabulation can be used to determine the existence, strength, direction, and pattern of an association.
- Elaboration analysis can be used in cross-tabular analysis to test for spurious and mediating relationships and to specify the conditions under which relationships occur.
- Inferential statistics are used with sample-based data to estimate the confidence that can be placed in a statistical estimate of a population parameter. Estimates of the probability that an association between variables may have occurred on the basis of chance are also based on inferential statistics.
- Regression analysis is a statistical method for characterizing the relationship between two or more quantitative variables with a linear equation and for summarizing the extent to which the linear equation represents that relationship. Correlation coefficients summarize the fit of the relationship to the regression line.

Discussion Questions

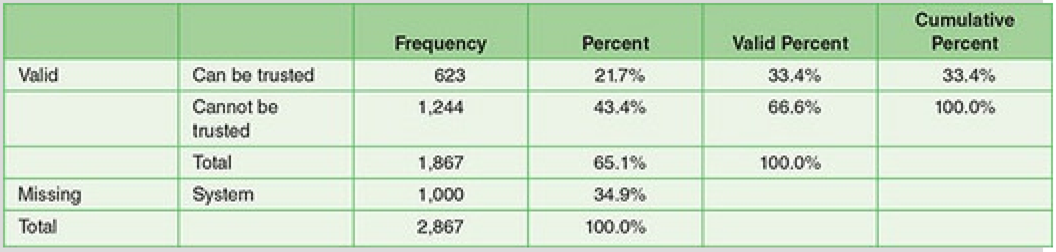
1. I presented in this chapter several examples of bivariate and trivariate cross-tabulations involving voting in the 2012 elections. What additional influences would you recommend examining to explain voting in elections? Suggest some additional independent variables for bivariate analyses with voting as well as several additional control variables to be used in three-variable crosstabs.
2. When should we control . . . just to be honest? In the evaluation project that I will describe in Chapter 13, I analyzed with some colleagues the effect on cognitive functioning of living in group homes rather than individual apartments. I found that living in group homes resulted in gains in cognitive functioning, compared with living in individual apartments. However, this benefit of group homes occurred only for residents who were not substance abusers; substance abusers did not gain cognitively from living in group (or individual) homes (Caplan et al. 2006). Would it have been all right if we had just reported the bivariate association between housing type and change in cognitive functioning? Should social researchers be expected to investigate alternative explanations for their findings? Should they be expected to check to see if the associations they find occur for different subgroups in their samples?

Practice Exercises

1. Exhibit 9.42 shows a frequency distribution of “trust in people” as produced by SPSS with the 2016 GSS data. As you can see, the table includes abbreviated labels for the variable and its response choices, as well as the raw frequencies and three percentage columns. The first percentage column (Percent) shows the percentage in each category of trust; the next percentage column (Valid Percent) is based on the total number of respondents who gave valid answers (1,867 in this instance). It is the Valid Percent column that normally should be used to construct a frequency distribution for presentation. The last percentage column is Cumulative Percent, adding up the valid percentages from top to bottom.

   Redo the table for presentation, using the format of the frequency distributions presented in the text (such as Exhibit 9.10).
2. Try your hand at recoding. Start with the distribution of the political ideology variable from Exhibit 9.10. It is named POLVIEWS in the GSS. Recode it to just three categories. What decision did you make about grouping? What was the consequence of this decision for the shape of the distribution? For the size of the middle category?
3. Cross-tabulations produced by most statistical packages are not in the proper format for inclusion in a report, and so they have to be reformatted. Referring to Exhibit 9.43, rewrite the table in presentational format, using one of the other tables as your guide. Describe the association in the table in terms of each of the four aspects of association. A chi-square test of statistical significance resulted in a *p* value of .000, meaning that the actual value was less than .001. State the level of confidence that you can have that the association in the table is not due to chance.

   Exhibit 9.42 Distribution of “Can People Be Trusted?”

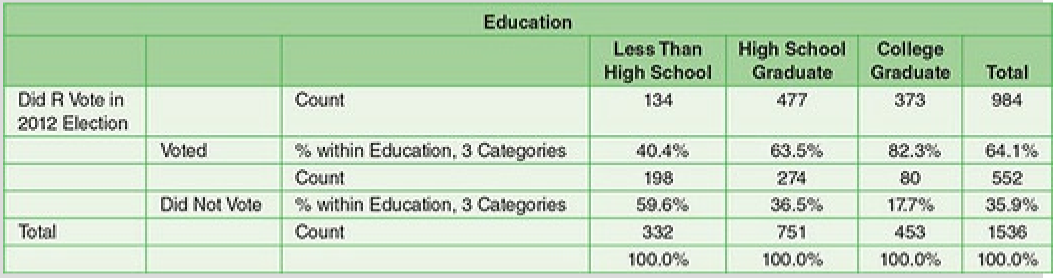
   

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Can be trusted	623	21.7%	33.4%	33.4%
	Cannot be trusted	1,244	43.4%	66.6%	100.0%
	Total	1,867	65.1%	100.0%	
Missing	System	1,000	34.9%		
Total		2,867	100.0%		

   Source: General Social Survey National Opinion Research Center 2016.

4. What if you had to answer this question: What was the income distribution of voters in the 2012 elections, and how did it compare with the income distribution for those who didn’t vote? Can you answer this question exactly with Exhibit 9.23? If not, change the column percentages in the table to row percentages. To do this, you will first have to convert the column percentages back to cell frequencies (although the frequencies are included in the table, so you can check your work). You can do this by multiplying the column percentage by the number of cases in the column, and then dividing by 100 (you will probably have fractional values because of rounding error). Then, compute the row percentage from these frequencies and the row totals.
5. Exhibit 9.43 contains a cross-tabulation of voting by education (recoded) directly as output by SPSS from the 2016 GSS data set. Describe the row and column marginal distributions. Try to calculate a cell percentage using the frequency (count) in that cell and the appropriate base number of cases.
6. Now, review the data analysis presented in the Cohen and Chaffee (2012) article on the book’s study site, at edge.sage pub.com/schutt9e, in which statistics were used. What do you learn from the data analysis? Which questions do you have about the meaning of the statistics used?
7. Test your understanding of basic statistics by completing one set of interactive exercises on quantitative data analysis from the study site.

Exhibit 9.43 Vote in 2012 Election by Education, in Three Categories



			Education			
			Less Than High School	High School Graduate	College Graduate	Total
Did R Vote in 2012 Election		Count	134	477	373	984
	Voted	% within Education, 3 Categories	40.4%	63.5%	82.3%	64.1%
		Count	198	274	80	552
	Did Not Vote	% within Education, 3 Categories	59.6%	36.5%	17.7%	35.9%
Total		Count	332	751	453	1536
			100.0%	100.0%	100.0%	100.0%

Source: General Social Survey National Opinion Research Center 2016.

Ethics Questions

1. Review the frequency distributions and graphs in this chapter. Change one of these data displays so that you are “lying with statistics.” (You might consider using the graphic technique discussed by Orcutt and Turner [1993].) How misleading is the resulting display?
2. Consider the relationship between voting and income that is presented in Exhibit 9.23. What third variables do you think should be controlled in the analysis to better understand the basis for this relationship? How might social policies be affected by finding out that this relationship was due to differences in neighborhood of residence rather than to income itself?

Web Exercises

1. Search the web for a social science example of statistics. Using the key terms from this chapter, describe the set of statistics you have identified. Which social phenomena does this set of statistics describe? What relationships, if any, do the statistics identify?
2. Go to the Roper Center for Public Opinion Research website at https://ropercenter.cornell.edu/. Now, pick the presidential approval ratings data, at https://presidential.roper.center/. Choose any two U.S. presidents from Franklin D. Roosevelt to the present. By using the website links, locate the presidential job performance poll data for the two presidents you have chosen.

   Based on poll data on presidential job performance, create a brief report that includes the following for each president you chose: the presidents you chose and their years in office; the question asked in the polls; and bar charts showing years when polls were taken, average of total percentage approving of job performance, average of total percentage disapproving of job performance, and average of total percentage with no opinion on job performance. Write a brief summary comparing and contrasting your two bar charts.
3. Do a web search for information on a social science subject in which you are interested. About what fraction of the information you find relies on statistics as a tool for understanding the subject? How do statistics allow researchers to test their ideas about the subject and generate new ideas? Write your findings in a brief report, referring to the websites you used.

Video Interview Questions

Listen to the interview with Peter Marsden for Chapter 9 at edge.sagepub.com/schutt9e.

1. What are the three goals of the General Social Survey (GSS)?
2. When was the first GSS conducted? Who developed the GSS concept?

SPSS Exercises

If you have been using the GSS2016x or GSS2016x_reduced data set for the SPSS exercises, you will now need to download the limited data sets designed for this chapter only: GSS2016y or GSS2016y_reduced (for the Student Version of SPSS). Please do this before proceeding. If you have been using the GSS2016 data set, you can just continue to use it.

1. Develop a description of the basic social and demographic characteristics of the U.S. population in 2016. Examine each characteristic with three statistical techniques: a graph, a frequency distribution, and a measure of central tendency (and a measure of variation, if appropriate).
    a. From the menu, select Graphs and then Legacy Dialogs and Bar. Select Simple Define [Marital—Category Axis]. Bars represent % of cases. Select Options (do not display groups defined by missing values). Finally, select Histogram for each of the variables [EDUC, EARNRS, INTWKDYH, ATTEND] from the GSS2016 data set.
    b. Describe the distribution of each variable.
    c. Generate frequency distributions and descriptive statistics for these variables. From the menu, select Analyze/Descriptive Statistics/Frequencies. In the Frequencies window, select MARITAL, EDUC, EARNRS, INTWKDYH, ATTEND, moving each to the Variable(s) window. Then choose Statistics, and select the mean, median, range, and standard deviation. Then choose Continue (and then OK).
    d. Collapse the categories for each distribution. Be sure to adhere to the guidelines given in the section "Grouped Data." Does the general shape of any of the distributions change as a result of changing the categories?
    e. Which statistics are appropriate to summarize the central tendency and variation of each variable? Do the values of any of these statistics surprise you?
2. Try describing relationships with support for capital punishment by using graphs. Select two relationships you identified in previous exercises and represent them in graphic form. Try drawing the graphs on lined paper (graph paper is preferable).
3. The GSS2016 data set you are using allows you to easily replicate the tables in this chapter. Try doing that. The computer output you get will probably not look like the tables shown here because I reformatted the tables for presentation, as you should do before preparing a final report. At this point, I'll let you figure out the menu commands required to generate these graphs, frequency distributions, and cross-tabulations. If you get flummoxed, review my instructions for the SPSS exercises in earlier chapters or go to the book study site and review my tutorials on SPSS.
4. Propose a variable that might have created a spurious relationship between income and voting. Explain your thinking. Propose a variable that might result in a conditional effect of income on voting, so that the relationship between income and voting would vary across the categories of the other variable. Test these propositions with three-variable cross-tabulations. Were any supported? How would you explain your findings?

Developing a Research Proposal

Use the GSS data to add a pilot study to your proposal (Exhibit 3.10, #18). A pilot study is a preliminary effort to test out the procedures and concepts that you have proposed to research.

1. In SPSS, review the GSS2016 variable list and identify some variables that have a connection to your research problem. If possible, identify one variable that might be treated as independent in your proposed research and one that might be treated as dependent.
2. Request frequencies for these variables.
3. Request a cross-tabulation of one dependent variable by a variable you are treating as independent. If necessary, recode the independent variable to five or fewer categories.

4. Write a brief description of your findings and comment on their implications for your proposed research. Did you learn any lessons from this exercise for your proposal?

Chapter 10 Qualitative Methods

- Research That Matters, Questions That Count
- Fundamentals of Qualitative Methods
    - History of Qualitative Research
    - Features of Qualitative Research
    - Basics of Qualitative Research
        - *The Case Study*
        - *Ethnography*
- Careers and Research
    - *Digital Ethnography*
- Participant Observation
    - Choosing a Role
        - *Covert Observation*
        - *Overt Observation*
        - *Overt Participation (Participant Observer)*
        - *Covert Participation*
- Research in the News: Family Life on Hold After Hurricane Harvey
    - Entering the Field
    - Developing and Maintaining Relationships
    - Sampling People and Events
    - Taking Notes
    - Managing the Personal Dimensions
- Intensive Interviewing
    - Establishing and Maintaining a Partnership
    - Asking Questions and Recording Answers
- Interviewing Online
- Focus Groups
- Generalizability in Qualitative Research
- Ethical Issues in Qualitative Research
- Conclusions

Hurricane Harvey roared ashore in southeastern Texas late on Friday, August 25, 2017 (National Oceanic and Atmospheric Administration [NOAA] 2017). The combined force of winds of up to 130 miles per hour, a storm surge, and heavy rain soon resulted in catastrophic and life-threatening flooding from the coast to Houston—the nation's fourth largest city—and beyond; by August 30, up to 50 inches of rain had fallen in some areas (see Exhibit 10.1). With initial estimates of the costs of the damage ranging above $100 billion, Harvey was one of the most devastating natural disasters in U.S. history, rivaled only by New Orleans's devastating Hurricane Katrina in 2005 (Quealy 2017). The resulting disruptions in individual lives and social patterns were equally profound. In the words of Kris Ford-Amofa, a mother of three who fled with her family from their new Houston home as the floodwaters began to seep into the second floor, her voice rising in frustration,

Research That Matters, Questions That Count

People can be very creative in trying to regain some social stability and meet basic needs after disaster strikes. After urban disasters, one form that this creativity can take is the development of urban gardening. Sociologist Yuki Kato at Tulane University in New Orleans and his collaborators Catarina Passidomo and Daina Harvey (2014) sought to understand how such efforts develop and the ways they are used as political tools for social transformation.

Using New Orleans after Hurricane Katrina as a case study, Kato, Passidomo, and Harvey conducted an ethnographic investigation of urban gardening projects using participant observation methods; the investigation had an intentional, political orientation of changing the allocation of resources. Their article describes four of the gardening projects they studied. They found that gardening projects ranged from the more political—"Our vision is to have the Lower Ninth Ward speak as one voice regarding what we want for food access in our neighborhood"—to the less political—"Hollygrove Market and Farm exists to increase accessibility of fresh produce to Hollygrove"—but their priorities and politics changed over time in relation to the broader political climate.

1. According to the authors, "we are careful to acknowledge that our analysis is situated within a particular spatial and temporal context; at the same time, however, we are interested in demonstrating themes and lessons offered by these projects that may be more or less universal in nature" (p. 2). In what ways do you think that a case study can be used to illuminate "universal" lessons?
2. What would you like to know about the particular methods used in this research to assess its authenticity?

In this chapter, you will learn the basic logic and procedures that guide qualitative research projects, as well as a bit about research on disasters. By the end of the chapter, you will understand the appeal of qualitative research methods and be able to discuss its strengths and limitations. After you read the chapter, you can enrich your understanding by reading the 2014 *Urban Studies* article by Kato, Passidomo, and Harvey at the *Investigating the Social World* study site and by completing the related interactive exercises for Chapter 10 at edge.sagepub.com/schutt9e.

Kato, Yuki, Catarina Passidomo, and Daina Harvey. 2014. "Political Gardening in Post-Disaster City: Lessons From New Orleans." *Urban Studies* 51:1833–1849.

Exhibit 10.1 Saving Pets in Hurricane Harvey

Source: Joe Raedle/Getty Images.

> I have no control over anything right now. I've never sat waiting for somebody to take care of me. I've always done it myself. Now, I have to wait all the time for somebody or something. I have to wait, wait, wait. (Healy 2017:76)

What could social research contribute to understanding Hurricane Harvey's impact and improving the response—both to this disaster, to the subsequent hurricanes Irma and Maria, and to those yet to come? The role of social researchers after Hurricane Katrina provides some examples. Within days after that hurricane, graduate students and then faculty researchers from the University of Delaware's Disaster Research Center (DRC) began to arrive in affected communities to study the storm's impact and the response to it (Thomas 2005). The research they designed used intensive interviews and participant observation as well as analysis of documents (Rodríguez, Trainor, and Quarantelli 2006). Other social researchers interviewed survivors and organized focus group discussions about the experience (Davis and Land 2007; Elder et al. 2007). In this chapter, I use this research on Hurricane Katrina and social science studies of other disasters to illustrate how sociologists learn by observing as they participate in a natural setting and to illustrate related qualitative methods. These examples will help you understand how some of our greatest insights into social processes can result from what appear to be very ordinary activities: observing, participating, listening, and talking.

But you will also learn that qualitative research is much more than just doing what comes naturally in social situations. Qualitative researchers must observe keenly, take notes systematically, question respondents strategically, and prepare to spend more time and invest more of their whole selves than often occurs with experiments or surveys. Moreover, if we are to have any confidence in a qualitative study's conclusions, each element of its design must be reviewed as carefully as we would review the elements of an experiment or survey. The result of careful use of these methods can also be insights into the features of the social world that are ill suited to investigation with experiments or surveys and to social processes that defy quantification.

The chapter begins with a brief history of qualitative methods and then an overview of the major features of qualitative research—using research on Hurricane Katrina and several other disasters for examples. The next section discusses the various approaches to participant observation research, which is the most distinctive qualitative method, and reviews the steps involved in participant observation. I then discuss the method of systematic observation, which is a quantitative approach to recording observational data; the contrast with participant observation will be instructive. In the following section, I review, in some detail, the issues involved in intensive interviewing before briefly explaining focus groups, an increasingly popular qualitative method. The last two sections discuss the challenge of generalizability in qualitative research and ethical issues that are of concern in any type of qualitative research project. By the chapter's end, you should appreciate the hard work required to translate "doing what comes naturally" into systematic research, be able to recognize strong and weak points in qualitative studies, and be ready to do some of it yourself.

Fundamentals of Qualitative Methods

Like quantitative methods, the term *qualitative methods* refers to a variety of research techniques that share some basic features. In this section, I will review the history of the development of these methods, identify the features they have in common, and provide more details about several specific qualitative methods that exemplify the approach.

History of Qualitative Research

Anthropologists and sociologists laid the foundation for modern qualitative methods while doing field research in the early decades of the 20th century. Dissatisfied with studies of native peoples that relied on secondhand accounts and inspection of artifacts, anthropologists Franz Boas and Bronislaw Malinowski went to live in or near the communities they studied. Boas visited Native American villages in the American Northwest; Malinowski lived among New Guinea natives. Neither truly participated in the ongoing social life of those they studied (Boas collected artifacts and original texts, and Malinowski reputedly lived as something of a noble among the natives he studied), but both helped establish the value of intimate familiarity with the community of interest and thus laid the basis for modern anthropology (Emerson 1983:2–5).

Many of sociology's field research pioneers were former social workers and reformers. Some brought their missionary concern with the spread of civic virtue among new immigrants to the Department of Sociology and Anthropology at the University of Chicago. Their successors continued to focus on the sources of community cohesion and urban strain but came to view the city as a social science "laboratory" rather than as a focus for reform. They adapted the fieldwork methods of anthropology to studying the "natural areas" of the city and the social life of small towns (Vidich and Lyman 2004). By the 1930s, 1940s, and 1950s, qualitative researchers were emphasizing the value of direct participation in community life and sharing in subjects' perceptions and interpretations of events (Emerson 1983:6–13). This naturalistic focus continued to dominate qualitative research into the 1960s, and qualitative researchers refined and formalized methods to develop the most realistic understanding of the natural social world (Denzin and Lincoln 2000:14–15).

The next two decades saw increasing emphasis on the way in which participants in the social world construct the reality that they experience and increasing disbelief that researchers could be disinterested observers of social reality or develop generalizable knowledge about it. Some adherents of the *constructivist* perspective urged qualitative researchers to describe particular events, rituals, and customs and to recognize that the interpretations they produced were no more "privileged" than were others' interpretations (Denzin and Lincoln 2000:15). "The making of every aspect of human existence is culturally created and determined in particular, localized circumstances about which no generalizations can be made. Even particularized meaning, however, is . . . relative and temporary" (Spretnak 1991:13–14).

In the 21st century, qualitative researchers are developing new techniques for studying social life online and for taking advantage of the increasing availability of pictures, videos, and texts on the Internet.

Field research: Research in which natural social processes are studied as they happen and left relatively undisturbed.

Features of Qualitative Research

Qualitative research designs share several features that distinguish them from experimental and survey research designs (Guba and Lincoln 1994; Maxwell 2005; Wolcott 1995):

Collection primarily of qualitative rather than quantitative data.

Any research design may collect both qualitative and quantitative data, but qualitative methods emphasize observations about natural behavior and artifacts that capture social life as participants experience it, rather than in categories the researcher predetermines. For example, the DRC researchers observed the response to the unprecedented threat posed by Hurricane Katrina in the major New Orleans hotels and concluded that the hotels went through three stages of "improvisation": (1) the hotels encouraged all guests who were not stranded to leave, (2) the hotel chains sent in food and other provisions for guests and staff, and (3) the hotels reorganized so that they could provide semipermanent lodging for federal disaster employees and evacuees (Rodríguez et al. 2006:87–89). The researchers described the different activities in the hotels at each of these "stages."

A focus on previously unstudied processes and unanticipated phenomena.

Previously unstudied attitudes and actions can't adequately be understood with a structured set of questions or within a highly controlled experiment. So qualitative methods have their greatest appeal when we need to explore new issues, investigate hard-to-study groups, or determine the meaning people give to their lives and actions. Disasters such as Hurricane Katrina certainly meet the criteria of unanticipated and hard to study. Dag Nordanger (2007:174) used qualitative methods to study loss and bereavement in war-torn Tigray, Ethiopia, because preliminary information indicated that people in this culture adjusted to loss in a very different way than do people in Western societies.

Exploratory research questions, with a commitment to inductive reasoning.

Qualitative researchers typically begin their projects seeking not to test preformulated hypotheses but to discover what people think, how they act, and why, in some social setting. The DRC researchers began their research by asking, "How did people, groups, and organizations in Louisiana react to the impact of Hurricane Katrina in September 2005?" (Rodríguez et al. 2006:83). Only after many observations do qualitative researchers try to develop general principles to account for their observations. It's still important to recognize that every researcher brings prior understandings and perspectives to their research; we can't fully erase our preconceptions and begin research as a *tabula rasa*—a blank slate. But we can do our best to look at the new setting we are studying with open eyes and a critical awareness of our own expectations. Which leads to the next feature.

Sensitivity to the subjective role of the researcher (*reflexivity*).

Qualitative researchers recognize that their perspective on social phenomena will reflect in part their own background and current situation. Who the researcher is and "where he or she is coming from" can affect what the research "finds." Some qualitative researchers believe that the goal of developing a purely "objective" view of the social world is impossible, but they discuss in their publications their own feelings about what they have studied so that others can consider how these feelings affected their findings. You can imagine how anthropology graduate student Hannah Gill (2004) had to consider her feelings when she encountered crime in the community she was studying in the Dominican Republic:

> On the second day I found myself flattened under a car to avoid getting shot by a woman seeking revenge for her husband's murder in the town market, and on the third I was sprinting away from a knife fight at a local hangout. (p. 2)

Rather than leaving, Gill assessed the danger, realized that although she stood out in the community as a "young, white American graduate student," she "was not a target and with necessary precautions, I would be relatively safe." She decided to use her experiences as "a constructive experience" that would give her insight into how people respond to risk.

Sociologist Barrie Thorne (1993) shared with readers her subjective reactions as she observed girls on a school playground:

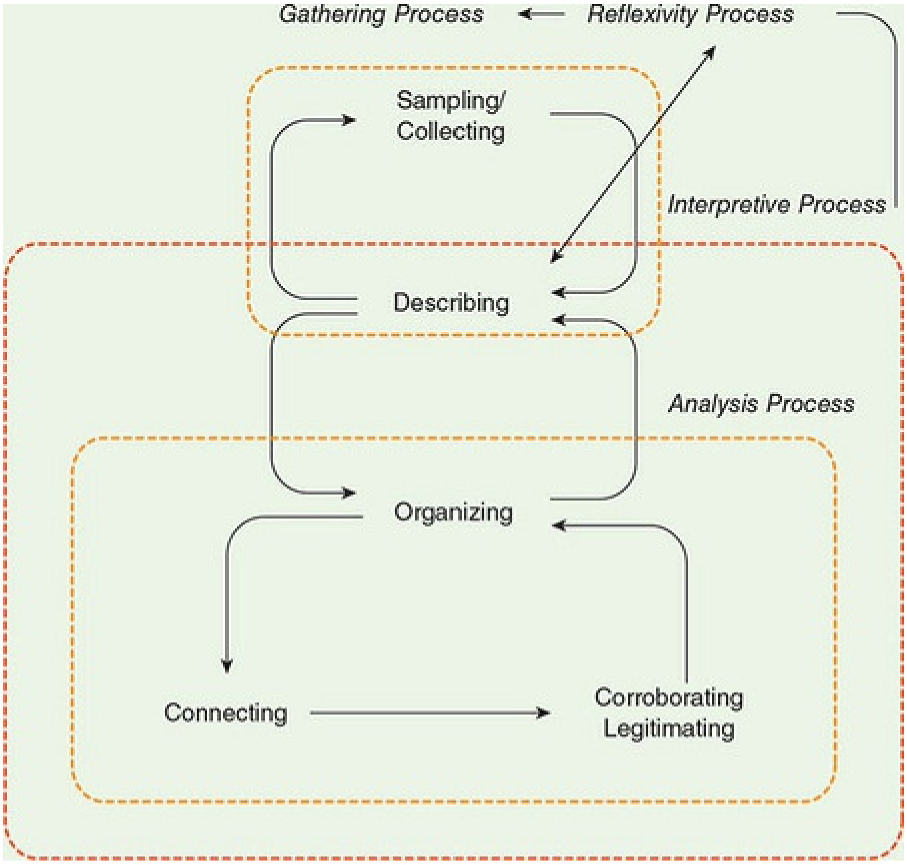
> I felt closer to the girls not only through memories of my own past, but also because I knew more about their gender-typed interactions. I had once played games like jump rope and statue buyer, but I had never ridden a skateboard and had barely tried sports like basketball and soccer. . . . Were my moments of remembering, the times when I felt like a ten-year-old girl, a source of distortion or insight? (p. 26)

Reflexivity: Sensitivity of and adaptation by the researcher to his or her influence in the research setting.

William Miller and Benjamin Crabtree (1999a) captured the entire process of qualitative research in a simple diagram (see Exhibit 10.2). In this diagram, qualitative research begins with the qualitative researcher reflecting on the setting and his or her relation to it and interpretations of it. The researcher then describes the goals and means for the research.

This description is followed by sampling and collecting data, describing the data, and organizing those data. Thus, the *gathering process* and the *analysis process* proceed together, with repeated description and analysis of data as they are collected and reflexive attention to the researcher's engagement in the process. As the data are organized, connections are identified between different data segments, and efforts are made to corroborate the credibility of these connections. This *interpretive process* begins to emerge in a written account that represents what has been done and how the data have been interpreted. Each of these steps in the research process informs the others and is repeated throughout the research process.

Exhibit 10.2 Qualitative Research Process


Gathering Process
Reflexivity Process
Sampling/ Collecting
Interpretive Process
Describing
Analysis Process
Organizing
Connecting
Corroborating Legitimating

Source: Miller and Crabtree (1999a:16). Reprinted with permission from SAGE Publications, Inc.

An orientation to social context, to the interconnections between social phenomena rather than to their discrete features.

The context of concern may be a program or organization, a community, or a broader social context. This feature of qualitative research is evident in Elif Kale-Lostuvali's (2007) description of Gölcük, a Turkish town, after the İzmit earthquake:

> For the first few months, the majority of the population lived in tents located either in tent cities or near partially damaged homes. Around mid-December, eligible survivors began to move into prefabricated houses built on empty hills around the center of the town. Many survivors had lost their jobs because of the earthquake. . . . Hence, daily life revolved around finding out about forms of provision. (p. 752)

A focus on human subjectivity, on the meanings that participants attach to events and that people give to their lives.

"Through life stories, people 'account for their lives.' . . . The themes people create are the means by which they interpret and evaluate their life experiences and attempt to integrate these experiences to form a self-concept" (Kaufman 1986:24–25). You can see this emphasis in an excerpt from an interview Nordanger (2007) conducted with a Tigrayan woman who had lost her property and her nine children in the preceding decades of war in Ethiopia:

> My name is the same. I was Mrs. NN, and now I am Mrs. NN. But I am not like the former. The former Mrs. NN had everything at hand, and was highly respected by others. People came to me for advice and help. But the recent Mrs. NN is considered like a half person. Though she does not go out for begging, she has the lifestyle of a beggar, so she is considered to be a beggar. (p. 179)

Use of idiographic rather than nomothetic causal explanation.

With its focus on particular actors and situations and the processes that connect them, qualitative research tends to identify causes as particular events embedded within an unfolding, interconnected action sequence (Maxwell 2005). The language of variables and hypotheses appears only rarely in the qualitative literature. Rodríguez et al. (2006) include in their analysis of "emergent and prosocial behavior following Hurricane Katrina" the following sequence of events in New Orleans hospitals:

> The floodwaters from the levee breaks created a new kind of crisis. Basements with stored food, water, and fuel, as well as morgues, were inundated. . . . As emergency generators ran out of fuel, the water, sewage, and air-conditioning systems failed. Patients who died in the hospitals had to be temporarily stored in stairwells. Eventually, waste of all kinds was strewn almost everywhere. The rising temperatures made most diagnostic equipment inoperable. . . . Regular hospital procedures simply stopped, but personnel improvised to try to provide at least minimum health care. For instance, physicians, nurses, and volunteers fanned

> patients to keep them cool, sometimes using manually operated devices to keep them breathing. (pp. 89–90)

Acceptance—by some qualitative researchers—of a constructivist philosophy.

Constructivist social scientists believe that social reality is socially constructed and that the goal of social scientists is to understand what meanings people give to reality, not to determine how reality works apart from these constructions. This philosophy rejects the positivist belief that there is a concrete, objective reality that scientific methods help us understand (Lynch and Bogen 1997); instead, constructivists believe that people construct an image of reality based on their own preferences and prejudices and their interactions with others and that this is as true of scientists as it is of everyone else in the social world. This means that we can never be sure that we have understood reality properly, that "objects and events are understood by different people differently, and those perceptions are the reality—or realities—that social science should focus on" (Rubin and Rubin 1995:35).

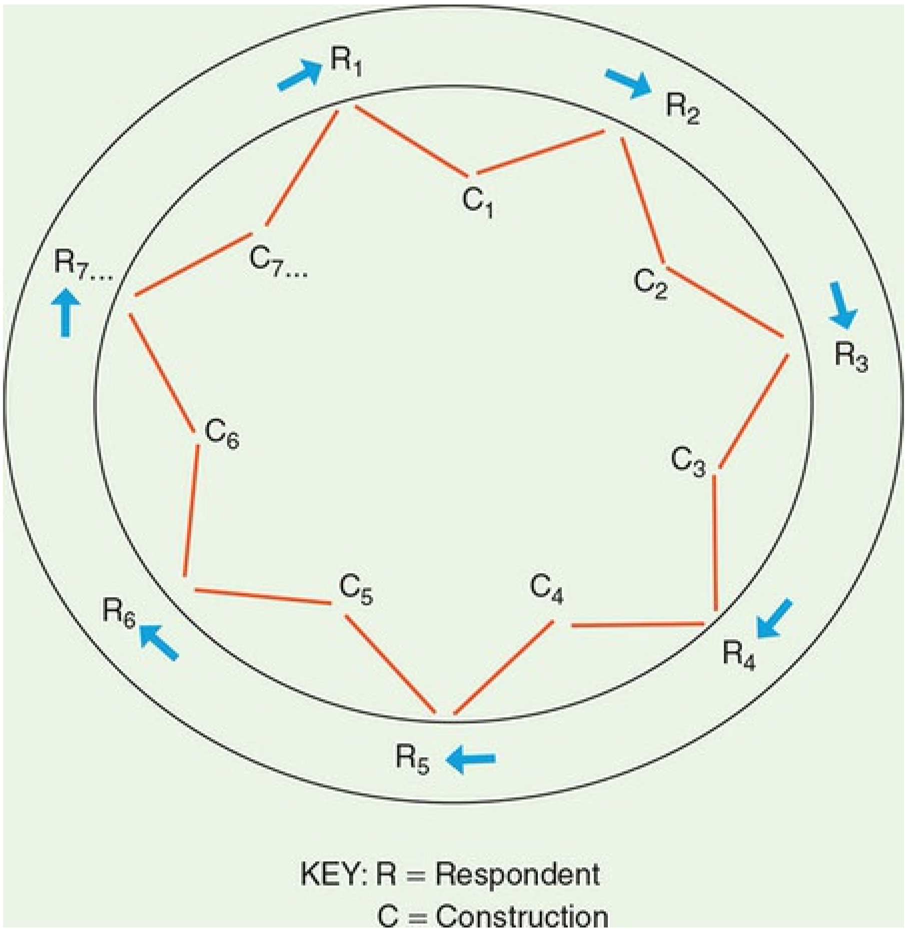
Constructivism emphasizes that different stakeholders in a social setting construct different beliefs (Guba and Lincoln 1989:44–45). Constructivists give particular attention to the different goals of researchers and other participants in a research setting and may seek to develop a consensus among participants about how to understand the focus of inquiry (Sulkunen 2008:73): "Truth is a matter of the best-informed and most sophisticated construction on which there is consensus at a given time" (Schwandt 1994:128).

Constructivist inquiry may use an interactive research process, in which a researcher begins an evaluation in some social setting by identifying the different interest groups in that setting. In a circular process known as a hermeneutic circle (Exhibit 10.3), the researcher interviews each respondent (R1, R2, etc.) to learn how they "construct" their thoughts and feelings about the topic of concern (C1, C2, etc.), and then gradually tries to develop a shared perspective on the problem being evaluated (Guba and Lincoln 1989:42, 180–181).

Constructivism: A methodology based on questioning belief in an external reality; emphasizes the importance of exploring the way in which different stakeholders in a social setting construct their beliefs.

Hermeneutic circle: A representation of the dialectical process in which the researcher obtains information from multiple stakeholders in a setting, refines his or her understanding of the setting, and then tests that understanding with successive respondents.

Exhibit 10.3 The Hermeneutic Circle



Source: Guba and Lincoln 1989. Fourth Generation Evaluation. SAGE.

Adaptive research design, in which the design develops as the research progresses.

> Each component of the design may need to be reconsidered or modified in response to new developments or to changes in some other component. . . . The activities of collecting and analyzing data, developing and modifying theory, elaborating or refocusing the research questions, and identifying and eliminating validity threats are usually all going on more or less simultaneously, each influencing all of the others. (Maxwell 2005:2–3)

Adaptive research design: A research design that develops as the research progresses.

You can see this adaptive quality in Kale-Lostuvali's (2007) description of his qualitative research work as he studied state–citizen encounters in the aftermath of the İzmit earthquake:

> I made my initial trip to Gölcük at the beginning of October 1999, six weeks after the earthquake. From then until the end of July 2000, I made two to three daylong trips per month, spending a total of 25 days in Gölcük. During these trips, I spent time mainly in the Gölcük Crisis Center, the administrative offices of two major tent cities, and two major prefab areas observing interactions. . . . As I got to know some of the state agents and survivors better, I began to hear their responses after specific interactions and their views of the provision and distribution process in general. In addition, I often walked around and spoke with many people in tent cities, in prefab areas, and in the center of the town. Sometimes, people I met in this way invited me to their homes and offices. (p. 752)

Ultimately, Kale-Lostuvali (2007:752) reported conversations with approximately 100 people, in-depth interviews with 30 carefully selected people, and many observational notes.

Basics of Qualitative Research

You can understand better how these different features make qualitative methods so distinct by learning the basics of specific qualitative methods and some of the insights those methods produced in leading studies. I will illustrate in this section the way in which qualitative research can produce insights about whole settings and cultures by presenting the basics of case study research and ethnographic research. I will also show how these approaches can be applied to research about social interaction on the Internet through the method of digital ethnography. I will then introduce the three qualitative methods that will be the focus of the rest of the chapter: participant observation, intensive interviewing, and focus groups.

The Case Study

Qualitative research projects often have the goal of developing an understanding of an entire slice of the social world, not just discrete parts of it. What was the larger social context in New Orleans after Hurricane Katrina (Rodríguez et al. 2006:87)? What was Chicago like during the 1995 Heat Wave, when thousands were hospitalized and more than 700 died of heat-related causes (Klinenberg 2002:1–9)? Sociologist Kai Erikson sent me the following verbal "picture" of New Orleans, as he observed it during a research trip a few days after Katrina:

> The carnage stretches out almost endlessly: more than a hundred thousand [crumpled] homes, at least fifty thousand [flattened] automobiles, the whole mass being covered by a crust of grey mud, dried as hard as fired clay by the sun. It was the silence of it, the emptiness of it; that is the story.

You can imagine the same reports after Hurricane Harvey: flooded neighborhoods, thousands of cars stranded on roadways, blocks of homes with water up to their roofs, people trying to make their way to shelter. It's not just a matters of "variables"—amounts of rain, numbers of deaths, property lost—but the feel of the whole situation—the "case."

Questions and images such as these reflect a concern with developing a case study. Case study is not so much a single method as it is a way of thinking about what a qualitative research project can, or perhaps should, focus on. The case may be an organization, a community, a social group, a family, or even an individual; as far as the qualitative researcher is concerned, it must be understood in its specific social context (Tight 2017:19–20). The idea is that the social world functions as an integrated whole; social researchers therefore need to develop "deep understanding of particular instances of phenomena" (Mabry 2008:214). By contrast, from this perspective, the quantitative research focus on variables and hypotheses mistakenly “slices and dices” reality in a way that obscures how the social world functions.

Case study: A setting or group that the analyst treats as an integrated social unit that must be studied holistically and in its particularity.

Educational researcher Robert Stake (1995) presents the logic of the case study approach thus:

> Case study is the study of the particularity and complexity of a single case, coming to understand its activity within important circumstances. . . . The qualitative researcher emphasizes episodes of nuance, the sequentiality of happenings in context, the wholeness of the individual. (pp. xi–xii)

Central to much qualitative case study research is the goal of creating a thick description of the setting studied—a description that provides a sense of what it is like to experience that setting from the standpoint of the natural actors in that setting (Geertz 1973). Stake’s (1995) description of “a case within a case,” a student in a school he studied, illustrates how a thick description gives a feel of the place and persons within it:

> At 8:30 a.m. on Thursday morning. Adam shows up at the cafeteria door. Breakfast is being served but Adam doesn’t go in. The woman giving out meal chits has her hands on him, seems to be sparring with him, verbally. And then he disappears. Adam is one of five siblings, all arrive at school in the morning with less than usual parent attention. Short, with a beautifully sculpted head . . . Adam is a person of notice.
>
> At 8:55 he climbs the stairs to the third floor with other upper graders, turning to block the girls behind them and thus a string of others. Adam manages to keep the girls off balance until Ms. Crain . . . spots him and gets traffic moving again. Mr. Garson . . . notices Adam, has a few quiet words with him before a paternal shove toward the room. (p. 150)

You will learn in the next sections how qualitative methodologists design research that can generate such thick descriptions of particular cases. I will also keep reminding you of the importance of “bounding the case” by defining carefully what the case is and the environment around it (Tight 2017:153).

Thick description: A rich description that conveys a sense of what it is like from the standpoint of the natural actors in that setting.

Ethnography

Ethnography is the study of a culture or cultures that a group of people share (Van Maanen 1995:4). Many qualitative researchers are guided by the tradition of ethnography, and most specific qualitative methods share some of its techniques. As a method, ethnography is usually meant to refer to the process by which a single investigator immerses himself or herself in a group for a long time (often one or more years), gradually establishing trust and experiencing the social world as do the participants (Madden 2010:16). Ethnographic research can be called *naturalistic,* because it seeks to describe and understand the natural social world as it is, in all its richness and detail. This goal is best achieved when an ethnographer is fluent in the local language and spends enough time immersed in the setting to know how people live, what they say about themselves and what they actually do, and what they value (Armstrong 2008:55; Fawcett and Pockett 2015:63–64).

Ethnography: The study of a culture or cultures that some group of people shares, using participant observation over an extended period.

Anthropological field research has traditionally been ethnographic, and much sociological fieldwork shares these same characteristics. But there are no particular methodological techniques associated with ethnography, other than just "being there." The analytic process relies on the thoroughness and insight of the researcher to "tell us like it is" in the setting, as he or she experienced it.

Code of the Street, Elijah Anderson's (1999) award-winning study of Philadelphia's inner city, captures the flavor of this approach:

> My primary aim in this work is to render ethnographically the social and cultural dynamics of the interpersonal violence that is currently undermining the quality of life of too many urban neighborhoods. . . . How do the people of the setting perceive their situation? What assumptions do they bring to their decision making? (pp. 10–11)

Like most traditional ethnographers, Anderson (1999) describes his concern with being "as objective as possible" and using his training as other ethnographers do, "to look for and to recognize underlying assumptions, their own and those of their subjects, and to try to override the former and uncover the latter" (p. 11). A rich description of life in the inner city emerges as Anderson's work develops. Although we often do not "hear" the residents speak, we feel the community's pain in Anderson's (1999) description of "the aftermath of death":

> When a young life is cut down, almost everyone goes into mourning. The first thing that happens is that a crowd gathers about the site of the shooting or the incident. The police then arrive, drawing more of a crowd. Since such a death often occurs close to the victim's house, his mother or his close relatives and friends may be on the scene of the killing. When they arrive, the women and girls often wail and moan, crying out their grief for all to hear, while the young men simply look on, in studied silence. . . . Soon the ambulance arrives. (p. 138)

Anderson (1999) uses this description as a foundation on which he develops key concepts, such as "code of the street":

> The "code of the street" is not the goal or product of any individual's action but is the fabric of everyday life, a vivid and pressing milieu within which all local residents must shape their personal routines, income strategies, and orientations to schooling, as well as their mating, parenting, and neighbor relations. (p. 326)

Anderson's report on his related Jelly's Bar study illustrates how his understanding deepened as he became more socially integrated into the group. He thus became more successful at "blending the local knowledge one has learned with what we already know sociologically about such settings" (Anderson 2003:236):

> I engaged the denizens of the corner and wrote detailed field notes about my experiences, and from time to time I looked for patterns and relationships in my notes. In this way, an understanding of the setting came to me in time, especially as I participated more fully in the life of the corner and wrote my field notes about my experiences; as my notes accumulated and as I reviewed them occasionally and supplemented them with conceptual memos to myself, their meanings became more clear, while even more questions emerged. (Anderson 2003:224)

A good ethnography like Anderson's is possible only when the ethnographer learns the subtleties of expression used in a group and the multiple meanings that can be given to statements or acts (Armstrong 2008:60–62). Good ethnographies also include some reflection by the researcher on the influence his or her own background has had on research plans, as well as on the impact of the research in the setting (Madden 2010:22–23).

Careers and Research

Dzenan Berberovic, Director of Development

Dzenan Berberovic was the first in his immediate family to attend college. While at the University of South Dakota, he earned a bachelor's degree in media and journalism with minors in communication studies and sociology.

During Berberovic's third year at the university, he was exposed to a research course. The use of research in marketing was eye-opening. It allowed him to see the important role of research in nearly every profession.

Berberovic's love for helping others, combined with his interest in both sociology and research, led him to pursue a career in the nonprofit sector. He now serves as the director of development for the University of South Dakota Foundation. Every day, he uses data and research completed on trends in the nonprofit and giving fields.

Berberovic's advice for students studying research methods is compelling: "Research is all around us. It will continue to grow, especially through the use of data analytics. Most professions will utilize a form of research; thus it is important to take advantage of the opportunities you are given as an undergraduate student. Even in careers like nonprofit—in my case—you may initially not think of research as a component of it. However, it plays a large role in moving organizations in the right direction."

Digital Ethnography

Communities can refer not only to people in a common physical location, but also to relationships that develop online. Online communities may be formed by persons with similar interests or backgrounds, perhaps to create new social relationships that location or schedules did not permit, or to supplement relationships that emerge in the course of work or school or other ongoing social activities. Like communities of people who interact face-to-face, online communities can develop a culture and become sources of identification and attachment (Kozinets 2010:14–15). And like physical communities, researchers can study online communities through immersion in the group for an extended period. Digital ethnography, also termed *netnography, cyberethnography* and *virtual ethnography* (James and Busher 2009:34–35), is the use of ethnographic methods to study online communities.

In some respects, digital ethnography is similar to traditional ethnography. The researcher prepares to enter the field by becoming familiar with online communities and their language and customs, formulating an exploratory research question about social processes or orientations in that setting, and selecting an appropriate community to study. Unlike in-person ethnographies, digital ethnographies can focus on communities whose members are physically distant and dispersed. The selected community should be relevant to the research question, involve frequent communication among actively engaged members, and have a number of participants who, as a result, generate a rich body of textual data (Kozinets 2010:89).

Digital ethnography: The use of ethnographic methods to study online communities; also termed *netnography, cyberethnography* and *virtual ethnography.*

The digital ethnographer's self-introduction should be clear and friendly. Robert Kozinets (2010:93) provides the following example written about the online discussion space alt.coffee:

> I've been lurking here for a while, studying online coffee culture on alt.coffee, learning a lot, and enjoying it very much . . . I just wanted to pop out of lurker status to let you know I am here . . . I will be wanting to quote some of the great posts that have appeared here, and I will contact the individuals by personal e-mail who posted them to ask their permission to quote them. I also will be making the document on coffee culture available to any interested members of the newsgroup for their perusal and comments—to make sure I get things right.

A digital ethnographer must keep both observational and reflective field notes but, unlike a traditional ethnographer, can return to review the original data—the posted text—long after it was produced. The data can then be coded, annotated with the researcher's interpretations, checked against new data to evaluate the persistence of social patterns, and used to develop a theory that is grounded in the data.

But are you feeling a bit uncomfortable about using the term *ethnography* to describe investigations of people interacting in cyberspace? If so, I'll bet you are thinking that you are missing important dimensions of people's feelings and their social interaction if you don't see their facial expressions, hear their intonations, or watch their body language. Not even an emoji can convey the rich sensory impressions we obtain in the course of face-to-face social interaction; interpersonal contact mediated by digital technology is not just the same as direct contact (Pink et al. 2016:3). Sherry Turkle (2015:7) puts it this way in her book about the need to "reclaim conversation":

> Computers offer the illusion of companionship without the demands of friendship and then, as the programs got really good, the illusion of friendship without the demands of intimacy. Because, face-to-face, people ask for things that computers never do. With people, things go best if you pay close attention and know how to put yourself in someone else's shoes. Real people demand responses to what they are feeling.

What's the take-home message? Much of the social world now happens online, and as social researchers we need to investigate it with the full array of methods that have proven useful in investigating other areas of the social world. But we always need to keep in mind the context of the larger social world in which the digital world develops, and consider the limitations we face if we try to understand the people who created the digital records only through their digital footprints (Pink et al. 2016:8–14).

It is now time to get into the specifics. The specifics of qualitative methods can best be understood by reviewing the three distinctive qualitative research techniques: participant observation, intensive (in-depth) interviewing, and focus groups. Participant observation and intensive interviewing are often used in the same project, whereas focus groups combine some elements of these two approaches into a unique data collection strategy. These techniques often can be used to enrich experiments and surveys. Qualitative methods can also be used in the study of textual or other documents as well as in historical and comparative research, but we will leave these research techniques for other chapters.

Participant observation: A qualitative method for gathering data that involves developing a sustained relationship with people while they go about their normal activities.

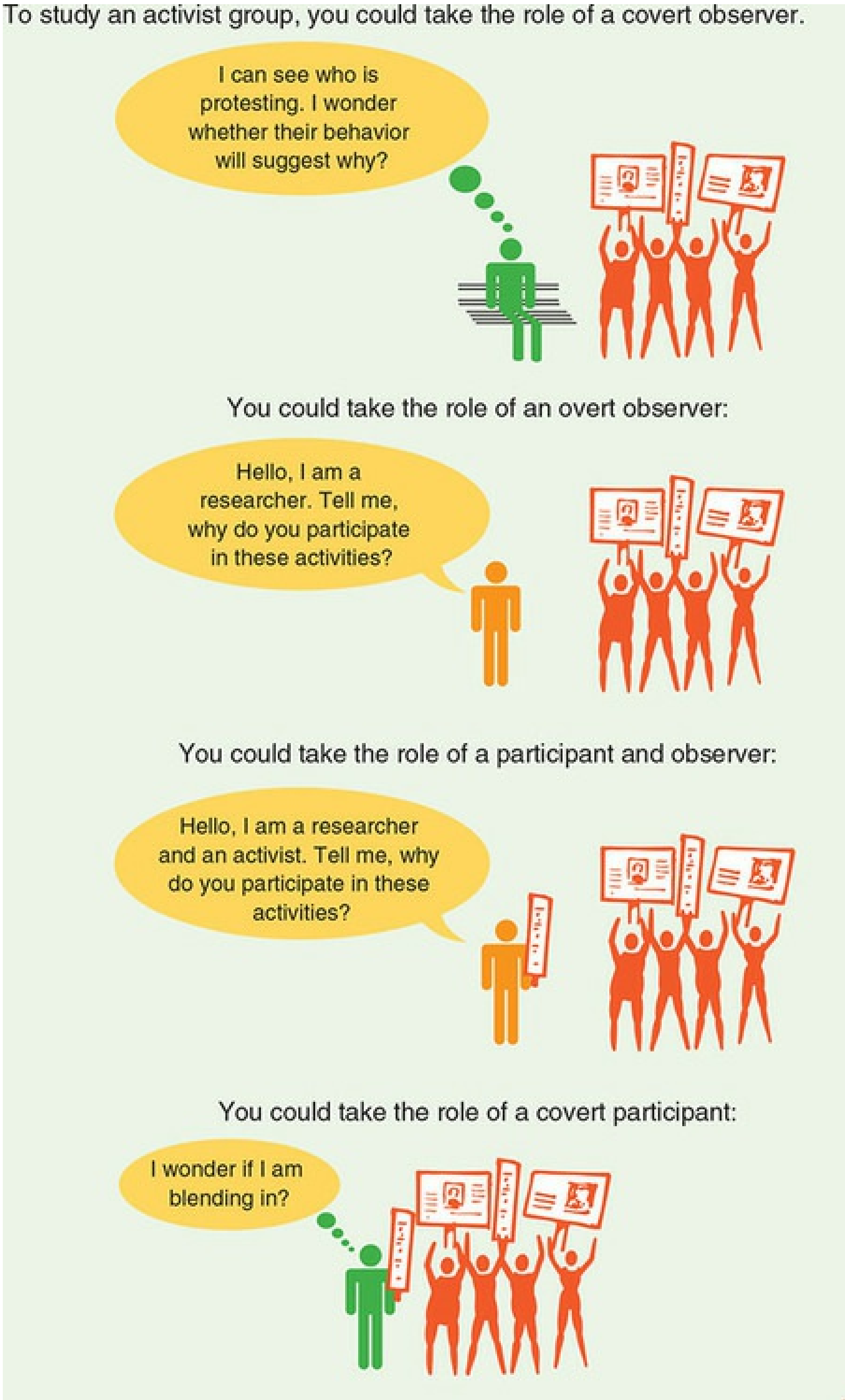
Intensive (in-depth) interviewing: A qualitative method that involves open-ended, relatively unstructured questioning in which the interviewer seeks in-depth information on the interviewee's feeling, experiences, and perceptions (Lofland and Lofland 1984:12).

Focus groups: A qualitative method that involves unstructured group interviews in which the focus group leader actively encourages discussion among participants on the topics of interest.

Participant Observation

Participant observation, termed *fieldwork* in anthropology and representing the core method of ethnographic research, was used by Rodríguez and his colleagues (2006) to study the aftermath of Hurricane Katrina, by Nordanger (2007) to study the effects of trauma in Ethiopia, and by Kale-Lostuvali (2007) to study the aftermath of the İzmit earthquake. Participant observation is a qualitative method in which natural social processes are studied as they happen (in "the field" rather than in the laboratory) and left relatively undisturbed. This is the classic field research method—a means for seeing the social world as the research subjects see it, in its totality, and for understanding subjects' interpretations of that world (Wolcott 1995:66). By observing people and interacting with them during their normal activities, participant observers seek to avoid the artificiality of experimental design and the unnatural structured questioning of survey research (Koegel 1987:8). This method encourages consideration of the context in which social interaction occurs, of the complex and interconnected nature of social relations, and of the sequencing of events (Bogdewic 1999:49).

Exhibit 10.4 The Participant Observation Continuum


To study an activist group, you could take the role of a covert observer.
I can see who is protesting. I wonder whether their behavior will suggest why?
You could take the role of an overt observer:
Hello, I am a researcher. Tell me, why do you participate in these activities?
You could take the role of a participant and observer:
Hello, I am a researcher and an activist. Tell me, why do you participate in these activities?
You could take the role of a covert participant:
I wonder if I am blending in?

The term *participant observer* actually refers to several different specific roles that a qualitative researcher can adopt (see Exhibit 10.4) (Gold 1958). As a covert observer, a researcher observes others without participating in social interaction and does not self-identify as a researcher.

This role is often adopted for studies in public places where there is nothing unusual about someone sitting and observing others. However, in many settings, a qualitative researcher will function as a complete observer, who does not participate in group activities and is publicly defined as a researcher. These two relatively passive roles contrast with the role of a researcher who participates actively in the setting. A qualitative researcher is a complete participant (also known as a *covert participant*) when she acts just like other group members and does not disclose her research role. If she publicly acknowledges being a researcher but nonetheless participates in group activities, she can be termed an *overt participant,* or true participant observer.

Covert observer: A role in participant observation in which the researcher does not participate in group activities and is not publicly defined as a researcher.

Complete (or overt) observer: A role in participant observation in which the researcher does not participate in group activities and is publicly defined as a researcher.

Complete (or covert) participant: A role in field research in which the researcher does not reveal his or her identity as a researcher to those who are observed while participating.

Participant observer: A researcher who gathers data through participating and observing in a setting where he or she develops a sustained relationship with people while they go about their normal activities. The term *participant observer* is often used to refer to a continuum of possible roles, from complete observation, in which the researcher does not participate along with others in group activities, to complete participation, in which the researcher participates without publicly acknowledging being an observer.

Choosing a Role

The first concern of every participant observer is to decide what balance to strike between observing and participating and whether to reveal one's role as a researcher. These decisions must take into account the specifics of the social situation being studied, the researcher's own background and personality, the larger sociopolitical context, and ethical concerns. Which balance of participating and observing is most appropriate also changes during most projects, and often many times. Moreover, the researcher's ability to maintain either a covert or an overt role will many times be challenged. Although the specifics differ in online research, digital ethnographers must also decide whether to announce their presence as researchers and how to participate in an online community, and then consider how these decisions affect the ongoing online social interaction they are observing (Hewson et al. 2016:37; Pink et al. 2016:13–14).

Covert Observation

In both observational roles, researchers try to see things as they happen, without actively participating in these events. Although there is no fixed formula to guide the observational process, observers try to identify the who, what, when, where, why, and how of the activities in the setting. Their observations will usually become more focused over time, as the observer develops a sense of the important categories of people and activities and gradually develops a theory that accounts for what is observed (Bogdewic 1999:54–56).

In social settings involving many people, in which observing while standing or sitting does not attract attention, covert observation is possible and is unlikely to have much effect on social processes. You may not even want to call this "covert" observation because your activities as an observer may be no different from those of others who are simply observing others to pass the time. However, when you take notes, when you systematically check out the different areas of a public space or different people in a crowd, when you arrive and leave at particular times to do your observing, you are acting differently in important respects from others in the setting. Moreover, when you write up what you have observed and, possibly, publish it, you have taken something unique from the people in that setting. If you adopt the role of a covert observer, you should always remember to evaluate how your actions in the setting and your purposes for being there may affect the actions of others and your own interpretations.

Overt Observation

When a researcher announces her role as a research observer, her presence is much more likely to alter the social situation being observed. This is the problem of reactive effects. It is not "natural" in most social situations for someone to be present who will record his or her observations for research and publication purposes, and so individuals may alter their behavior. The overt, or complete, observer is even more likely to have an impact when the social setting involves few people or if observing is unlike the usual activities in the setting. Observable differences between the observer and those being observed also increase the likelihood of reactive effects. For example, some children observed in the research by Thorne (1993:16–17) treated her as a teacher when she was observing them in a school playground and so asked her to resolve disputes. No matter how much she tried to remain aloof, she still appeared to children as an adult authority figure and so experienced pressure to participate (Thorne 1993:20). However, in most situations, even overt observers find that their presence seems to be ignored by participants after a while and to have no discernible impact on social processes.

Overt Participation (Participant Observer)

Most field researchers adopt a role that involves some active participation in the setting. Usually, they inform at least some group members of their research interests, but then they participate in enough group activities to develop rapport with members and to gain a direct sense of what group members experience. This is not an easy balancing act.

> The key to participant observation as a fieldwork strategy is to take seriously the challenge it poses to participate more, and to play the role of the aloof observer less. Do not think of yourself as someone who needs to wear a white lab coat and carry a clipboard to learn about how humans go about their everyday lives. (Wolcott 1995:100)

Field researcher: A researcher who uses qualitative methods to conduct research in the field.

Reactive effects: The changes in individual or group behavior that result from being observed or otherwise studied.

Nordanger (2007) described how, accompanied by a knowledgeable Tigrayan research assistant, he developed rapport with community members in Tigray, Ethiopia:

> Much time was spent at places where people gathered, such as markets, “sewa houses” (houses for homebrewed millet beer: sewa), cafés, and bars, and for the entire study period most invitations for a drink or to go to people’s homes for *injerra* (the sour pancake that is their staple food) and coffee ceremonies were welcomed. The fact that the research topic garnered interest and engagement

> made access to relevant information easy. Numerous informal interviews derived from these settings, where the researcher discussed his interests with people. (p. 176)

Participating and observing have two clear ethical advantages as well. Because group members know the researcher's real role in the group, they can choose to keep some information or attitudes hidden. By the same token, the researcher can decline to participate in unethical or dangerous activities without fear of exposing his or her identity.

Most field researchers who opt for disclosure get the feeling that, after they have become known and at least somewhat trusted figures in the group, their presence does not have any palpable effect on members' actions. The major influences on individual actions and attitudes are past experiences, personality, group structure, and so on, so the argument goes, and these continue to exert their influence even when an outside observer is present. The participant observer can then be ethical about identity disclosure and still observe the natural social world. In practice, however, it can be difficult to maintain a fully open research role in a setting in which new people come and go, often without providing appropriate occasions during which the researcher can disclose his or her identity.

Of course, the argument that the researcher's role can be disclosed without affecting the social process under investigation is less persuasive when the behavior to be observed is illegal or stigmatized, so that participants have reasons to fear the consequences of disclosure to any outsider. Konstantin Belousov and his colleagues (2007) provide a dramatic example of this problem from their fieldwork on regulatory enforcement in the Russian shipping industry. In a setting normally closed to outsiders and linked to organized crime, the permission of a port official was required. However, this official was murdered shortly after the research began. After that "our presence was now barely tolerated, and to be avoided at all costs. . . . explanations became short and respondents clearly wished to get rid of us as soon as possible" (pp. 164–165).

Even when researchers maintain a public identity as researchers, ethical dilemmas arising from participation in the group activities do not go away. In fact, researchers may have to "prove themselves" to the group members by joining in some of their questionable activities. For example, police officers gave John Van Maanen (1982) a nonstandard and technically prohibited pistol to carry on police patrols. Harold Pepinsky (1980) witnessed police harassment of a citizen but did not intervene when the citizen was arrested. Trying to strengthen his ties with a local political figure in his study of a poor Boston community he called Cornerville, William Foote Whyte (1955) illegally voted multiple times in a local election.

Experienced participant observers try to lessen some of the problems of identity disclosure by evaluating both their effect on others in the setting and the effect of others on the observers writing about these effects throughout the time they are in the field and while they analyze their data. They also are sure, while in the field, to preserve some physical space and regular time when they can concentrate on their research and schedule occasional meetings with other researchers to review the fieldwork. Participant observers modify their role as circumstances seem to require, perhaps not always disclosing their research role at casual social gatherings or group outings, but being sure to inform new members of it.

Covert Participation

To lessen the potential for reactive effects and to gain entry to otherwise inaccessible settings, some field researchers have adopted the role of covert participants, keeping their research secret and trying their best to act similar to other participants in a social setting or group. Laud Humphreys (1970) took the role of a covert participant when he served as a "watch queen" so that he could learn about the men engaging in homosexual acts in a public restroom. Randall Alfred (1976) joined a group of Satanists to investigate the group members and their interaction. Erving Goffman (1961) worked as a state hospital assistant while studying the treatment of psychiatric patients.

In the News Research in the News: Family Life on Hold After Hurricane Harvey

For Further Thought?

Brown water "slithered under the front door" of her first home before it began "crawling up the stairs toward the bedrooms" on the second floor where she thought they would be safe; then it "swirled around her three children" as they waded up the street to safety. Kris Ford-Amofa had a lot to worry about and, after having received no response to her pleas on Facebook—"We need a boat asap!!!"—and failing to find an online form or get through to the right person at FEMA, and returning days later to their "American dream" of a home with its now buckled living room floor and collapsing walls, and a never ending to-do list, she and her husband knew that "things are not the way they used to be."

A journalist's story can provide some of the richness of a qualitative case study and the insights that come from ethnographic immersion in a social context, but it's not the same as these and other systematic approaches to social research.

1. How well do you understand the social context of the disaster experience from a story like this? What else would like to know? What questions would you like to ask survivors and how would you select them?
2. What opportunities for conducting an ethnographic investigation can you think of in a disaster and recovery situation like Hurricane Harvey? What would be some of the problems, even if you were living there and were familiar with the city?

News source: Healy, Jack. 2017. "For One Family in Houston, an Overwhelming Start to Recovery." *The New York Times,* September 3, pp. A1, A17.

Although the role of a covert participant lessens some of the reactive effects encountered by the complete observer, covert participants confront other problems:

- *Covert participants cannot take notes openly or use any obvious recording devices.* They must write up notes based solely on their memory and must do so at times when it is natural for them to be away from the group members.
- *Covert participants cannot ask questions that will arouse suspicion.* Thus, they often have trouble clarifying the meaning of other participants' attitudes or actions.
- *The role of a covert participant is difficult to play successfully.* Covert participants will not know how the regular participants would act in every situation in which the

researchers find themselves. Regular participants have entered the situation from different social backgrounds and with goals different from that of the researchers. Researchers' spontaneous reactions to every event are unlikely to be consistent with those of the regular participants (Mitchell 1993). Suspicion that researchers are not "one of us" may then have reactive effects, obviating the value of complete participation (Erikson 1967). In his study of the Satanists, for example, Alfred (1976) pretended to be a regular group participant until he completed his research, at which time he informed the group leader of his covert role. Rather than act surprised, the leader told Alfred that he had long considered Alfred to be "strange," not similar to the other people—and we will never know for sure how Alfred's observations were affected.

- *Covert participants need to keep up the act at all times while in the setting under study.* Researchers may experience enormous psychological strain, particularly in situations where they are expected to choose sides in intragroup conflict or to participate in criminal or other acts. Of course, some covert observers may become so wrapped up in the role they are playing that they adopt not only just the mannerisms but also the perspectives and goals of the regular participants—that is, they "go native." At this point, they abandon research goals and cease to evaluate critically what they are observing.

Ethical issues have been at the forefront of debate over the strategy of covert participation. Erikson (1967) argued that covert participation is, by its very nature, unethical and should not be allowed except in public settings. Erikson points out that covert researchers cannot anticipate the unintended consequences of their actions for research subjects. If other people suspect the identity of the researcher or if the researcher contributes to or impedes group action, the consequences can be adverse. In addition, other social scientists are harmed either when covert research is disclosed during the research or on its publication because distrust of social scientists increases and access to research opportunities may decrease.

However, a total ban on covert participation would "kill many a project stone dead" (Punch 1994:90). Studies of unusual religious or sexual practices and of institutional malpractice would rarely be possible. "The crux of the matter is that some deception, passive or active, enables you to get at data not obtainable by other means" (Punch 1994:91). Richard Mitchell Jr. (1993) presents the argument of some researchers that the social world "is presumed to be shot through with misinformation, evasion, lies, and fronts at every level, and research in kind—secret, covert, concealed, and disguised—is necessary and appropriate" (p. 30). Therefore, some field researchers argue that covert participation is legitimate in some of the settings. If the researcher maintains the confidentiality of others, keeps commitments to others, and does not directly lie to others, some degree of deception may be justified in exchange for the knowledge gained (Punch 1994:90).

Entering the Field

Entering the field, the setting under investigation, is a critical stage in a participant observation project because it can shape many subsequent experiences. Some background work is necessary before entering the field—at least enough to develop a clear understanding of what the research questions are likely to be and to review one's personal stance toward the people and problems likely to be encountered. You need to have a sense of the social boundaries around the setting or "case" you will study. With participant observation, researchers must also learn in advance how participants dress and what their typical activities are to avoid being caught completely unaware. Finding a participant who can make introductions is often critical (Rossman and Rallis 1998:102–103), and formal permission may be needed in an organization's setting (Bogdewic 1999:51–53). It may take weeks or even months before entry is possible.

Timothy Diamond (1992) applied to work as an assistant to conduct research as a participant observer in a nursing home. His first effort failed miserably:

> My first job interview. . . . The administrator of the home had agreed to see me on [the recommendation of two current assistants]. The administrator . . . probed suspiciously, "Now why would a white guy want to work for these kinds of wages?" . . . He continued without pause, "Besides, I couldn't hire you if I wanted to. You're not certified." That, he quickly concluded, was the end of our interview, and he showed me to the door. (pp. 8–9)

After taking a course and receiving his certificate, Diamond was able to enter the role of nursing assistant as others did.

Many field researchers avoid systematic study and extensive reading about a setting for fear that it will bias their first impressions, but entering without any sense of the social norms can lead to a disaster. Whyte came close to such a disaster when he despaired of making any social contacts in Cornerville and decided to try an unconventional entry approach (i.e., unconventional for a field researcher). In *Street Corner Society,* Whyte (1955) describes what happened when he went to a hotel bar in search of women to talk to:

> I looked around me again and now noticed a threesome: one man and two women. It occurred to me that here was a maldistribution of females which I might be able to rectify. I approached the group and opened with something like this: "Pardon me. Would you mind if I joined you?" There was a moment of silence while the man stared at me. He then offered to throw me downstairs. I

> assured him that this would not be necessary and demonstrated as much by walking right out of there without any assistance. (p. 289)

Whyte needed a gatekeeper who could grant him access to the setting; he finally found one in “Doc” (Rossman and Rallis 1998:108–111). A helpful social worker at the local settlement house introduced Whyte to this respected leader, who agreed to help:

> Well, any nights you want to see anything, I’ll take you around. I can take you to the joints—gambling joints—I can take you around to the street corners. Just remember that you’re my friend. That’s all they need to know [so they won’t bother you]. (Whyte 1955:291)

You have already learned that Nordanger (2007:176) relied on a gatekeeper to help him gain access to local people in Tigray, Ethiopia.

When participant observing involves public figures who are used to reporters and researchers, a more direct approach may secure entry into the field. Richard Fenno (1978:257) used this direct approach in his study of members of the U.S. Congress: He simply wrote and asked permission to observe selected members of the Congress at work. He received only two refusals, attributing this high rate of subject cooperation to such reasons as interest in a change in the daily routine, commitment to making themselves available, a desire for more publicity, the flattery of scholarly attention, and interest in helping teach others about politics. Other groups have other motivations, but in every case, some consideration of these potential motives in advance should help smooth entry into the field.

In short, field researchers must be very sensitive to the impression they make and to the ties they establish when entering the field. This stage lays the groundwork for collecting data from people who have different perspectives and for developing relationships that the researcher can use to surmount the problems in data collection that inevitably arise in the field. The researcher should be ready with a rationale for his or her participation and some sense of the potential benefits to participants. Discussion about these issues with key participants or gatekeepers should be honest and identify what the participants can expect from the research, without necessarily going into detail about the researcher’s hypotheses or research questions (Rossman and Rallis 1998:51–53, 105–108).

Gatekeeper: A person in a field setting who can grant researchers access to the setting.

Developing and Maintaining Relationships

Researchers must be careful to manage their relationships in the research setting so that they can continue to observe and interview diverse members of the social setting throughout the long period typical of participant observation (Maxwell 2005:82–87). Every action the researcher takes can develop or undermine this relationship. Interaction early in the research process is particularly sensitive because participants don't know the researcher and the researcher doesn't know the routines. Thorne (1993) felt she had gained access to kids' more private world "when kids violated rules in my presence, such as swearing or openly blowing bubble gum where these acts were forbidden, or swapping stories about recent acts of shoplifting" (pp. 18–19). Conversely, Van Maanen (1982) found his relationship with police officers undermined by one incident:

> Following a family beef call in what was tagged the Little Africa section of town, I once got into what I regarded as a soft but nonetheless heated debate with the officer I was working with that evening on the merits of residential desegregation. My more or less liberal leanings on the matter were bothersome to this officer, who later reported my disturbing thoughts to his friends in the squad. Before long, I was an anathema to this friendship clique and labeled by them undesirable. Members of this group refused to work with me again. (p. 110)

So Van Maanen failed to maintain a research (or personal) relationship with this group. Do you think he should have kept his opinions about residential desegregation to himself? How honest should field researchers be about their feelings? Should they "go along to get along"?

Whyte used what, in retrospect, was a sophisticated two-part strategy to develop and maintain a relationship with the Cornerville street-corner men. The first part of Whyte's strategy was to maintain good relations with Doc and, through Doc, to stay on good terms with the others. Doc became a key informant in the research setting—a knowledgeable insider who knew the group's culture and was willing to share access and insights with the researcher (Gilchrist and Williams 1999). The less obvious part of Whyte's strategy was a consequence of his decision to move into Cornerville, a move he decided was necessary to really understand and be accepted in the community. The room he rented in a local family's home became his base of operations. In some respects, this family became an important dimension of Whyte's immersion in the community: He tried to learn Italian by speaking with the family members, and they conversed late at night as if Whyte were a real family member. But Whyte recognized that he needed a place to unwind after his days of constant alertness in the field, so he made a conscious decision not to include the family as an object of study. Living in this family's home became a means for Whyte to maintain standing as a community insider without becoming totally immersed in the demands of research (Whyte 1955:294–297).

Key informant: An insider who is willing and able to provide a field researcher with superior access and information, including answers to questions that arise in the course of the research.

Experienced participant observers have developed some sound advice for others seeking to maintain relationships in the field (Bogdewic 1999:53–54; Rossman and Rallis 1998:105–108; Whyte 1955:300–306; Wolcott 1995:91–95):

- Develop a plausible (and honest) explanation for yourself and your study.
- Maintain the support of key individuals in groups or organizations under study.
- Be unobtrusive and unassuming. Don't "show off" your expertise.
- Don't be too aggressive in questioning others (e.g., don't violate implicit norms that preclude discussion of illegal activity with outsiders). Being a researcher requires that you do not simultaneously try to be the guardian of law and order. Instead, be a reflective listener.
- Ask very sensitive questions only of informants with whom your relationship is good.
- Be self-revealing, but only up to a point. Let participants learn about you as a person, but without making too much of yourself.
- Don't fake your social similarity with your subjects. Taking a friendly interest in them should be an adequate basis for developing trust.
- Avoid giving or receiving monetary or other tangible gifts but without violating norms of reciprocity. Living with other people, taking others' time for conversations, and going out for a social evening all create expectations and incur social obligations, and you can't be an active participant without occasionally helping others. But you will lose your ability to function as a researcher if you come to be seen as someone who gives away money or other favors. Such small forms of assistance as an occasional ride to the store or advice on applying to college may strike the right balance.
- Be prepared for special difficulties and tensions if multiple groups are involved. It is hard to avoid taking sides or being used in situations of intergroup conflict.

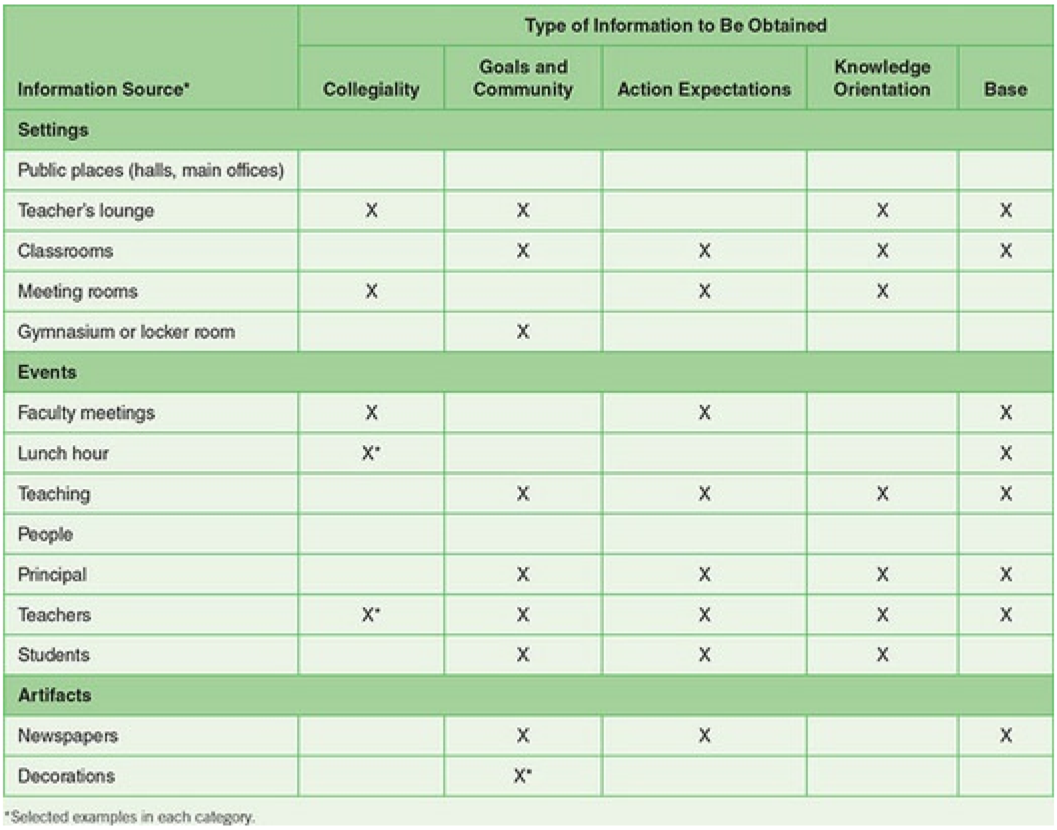
Sampling People and Events

In qualitative research, the need to intensively study the people, places, or phenomena of interest guide sampling decisions. Most qualitative researchers limit their focus to just one or a few sites or programs, so that they can focus all their attention on the social dynamics of those settings. This focus on a limited number of cases does not mean that sampling is unimportant. The researcher must be reasonably confident about gaining access and that the site can provide relevant information. The sample must be appropriate and adequate for the study, even if it is not representative. The qualitative researcher may select a *critical case* that is unusually rich in information pertaining to the research question, a *typical case* precisely because it is judged to be typical, or a *deviant case* that provides a useful contrast (Kuzel 1999). Within a research site, plans may be made to sample different settings, people, events, and artifacts (see Exhibit 10.5).

Studying more than one case or setting almost always strengthens the causal conclusions and makes the findings more generalizable (King, Keohane, and Verba 1994). The DRC researchers (Rodríguez et al. 2006:87) studied emergent behavior in five social "groupings": hotels, hospitals, neighborhood groups, rescue teams, and the Joint Field Office (JFO). To make his conclusions more generalizable, Diamond (1992:5) worked in three different Chicago nursing homes "in widely different neighborhoods" that had very different proportions of residents supported by Medicaid. He then "visited many homes across the United States to validate my observations" (p. 5). Klinenberg (2002:79–128) contrasted the social relations in two Chicago neighborhoods. Thorne (1993:6–7) observed in a public elementary school in California for 8 months and then, 4 years later, in a public elementary school in Michigan for 3 months.

Other approaches to sampling in field research are more systematic. You have already learned in Chapter 5 about some of the nonprobability sampling methods used in field research. For instance, purposive sampling can be used to identify the opinion of the leaders and representatives of different roles. With snowball sampling, field researchers learn from participants about who represents different subgroups in a setting. Quota sampling also may be employed to ensure the representation of particular categories of participants. Using some type of intentional sampling strategy within a particular setting can allow tests of some hypotheses, which would otherwise have to wait until comparative data could be collected from several other settings (King et al. 1994).

Exhibit 10.5 Sampling Plan for a Participant Observation Project in Schools



Information Source*	Type of Information to Be Obtained				
	Collegiality	Goals and Community	Action Expectations	Knowledge Orientation	Base
Settings					
Public places (halls, main offices)					
Teacher's lounge	X	X		X	X
Classrooms		X	X	X	X
Meeting rooms	X		X	X	
Gymnasium or locker room		X			
Events					
Faculty meetings	X		X		X
Lunch hour	X*				X
Teaching		X	X	X	X
People					
Principal		X	X	X	X
Teachers	X*	X	X	X	X
Students		X	X	X	
Artifacts					
Newspapers		X	X		X
Decorations		X*			

*Selected examples in each category.

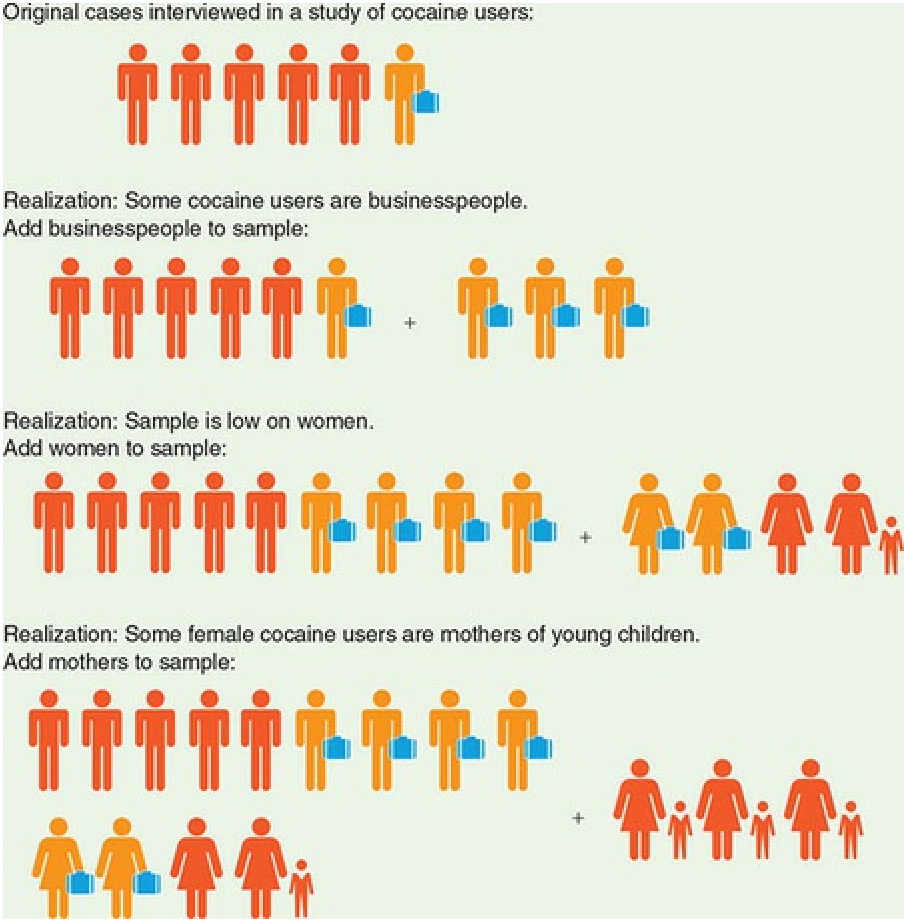
Source: Adapted from Marshall and Rossman (2011:108–110). Reprinted with permission from SAGE Publications, Inc.

Theoretical sampling is a systematic approach to sampling in participant observation studies (Glaser and Strauss 1967). When field researchers discover in an investigation that particular processes seem to be important, inferring that certain comparisons should be made or that similar instances should be checked, the researchers then choose new settings or individuals that permit these comparisons or checks (Ragin 1994:98–101) (see Exhibit 10.6). Spencer Moore and colleagues' (2004) strategy for selecting key informants in their research on North Carolina's Hurricane Floyd experience exemplifies this type of approach:

> Sixteen key informant interviews were conducted with volunteers for local nonprofit organizations, community and religious leaders, and local government officials in all five HWC-project counties [Health Works in the Community]. These representatives were chosen on the basis of their county or city administrative position (e.g., emergency management or assistant county managers), as well as on the basis of leadership in flood-related relief activities, as identified by local officials or as reported in local newspapers. (p. 209)

Theoretical sampling: A sampling method recommended for field researchers by Glaser and Strauss (1967). A theoretical sample is drawn in a sequential fashion, with settings or individuals selected for study as earlier observations or interviews indicate that these settings or individuals are influential.

Exhibit 10.6 Theoretical Sampling


Original cases interviewed in a study of cocaine users:
Realization: Some cocaine users are businesspeople.
Add businesspeople to sample:
+
Realization: Sample is low on women.
Add women to sample:
+
Realization: Some female cocaine users are mothers of young children.
Add mothers to sample:
+

When field studies do not require ongoing, intensive involvement by researchers in the setting, the experience sampling method (ESM) can be used. In this method, the experiences, thoughts, and feelings of a number of people are sampled randomly as they go about their daily activities. Participants in an ESM study carry an electronic pager and fill out reports when they are beeped. For example, 107 adults carried pagers in Robert Kubey's (1990) ESM study of television habits and family quality of life. Participants' reports indicated that heavy TV viewers were less active during non-TV family activities, although heavy TV viewers also spent more time with their families and felt as positively toward other family members as did those who watched less TV.

Experience sampling method (ESM): A technique for drawing a representative sample of everyday activities, thoughts, and experiences. Participants carry a pager and are beeped at random times over several days or weeks; on hearing the beep, participants complete a report designed by the researcher.

Although ESM is a powerful tool for field research, it is still limited by the need to recruit people to carry pagers. Ultimately, the generalizability of ESM findings relies on the representativeness, and reliability, of the persons who cooperate in the research.

Taking Notes

Written notes are the primary means of recording participant observation data (Emerson, Fretz, and Shaw 1995). Of course, “written” no longer means only handwritten; many field researchers jot down partial notes while observing and then retreat to their computers to write up more complete notes on a daily basis. The computerized text can then be inspected and organized after it is printed out, or it can be marked up and organized for analysis using one of several computer programs designed especially for the task.

It is almost always a mistake to try to take comprehensive notes while engaged in the field —the process of writing extensively is just too disruptive. The usual procedure is to jot down brief notes about highlights of the observation period. These brief notes, called jottings, can then serve as memory joggers when the researcher is writing the actual field notes at a later session. It will also help maintain a daily log in which each day’s activities are recorded (Bogdewic 1999:58–67). With the aid of the jottings and some practice, researchers usually remember a great deal of what happened—as long as the comprehensive field notes are written immediately afterward, or at least within the next 24 hours, and before they have discussed them with anyone else.

Jottings: Brief notes written in the field about highlights of an observation period.

Field notes: Notes that describe what has been observed, heard, or otherwise experienced in a participant observation study. These notes usually are written after the observational session.

The following excerpts shed light on the note-taking processes that Diamond and Thorne used while in the field. Taking notes was more of a challenge for Diamond (1992) because many people in the setting did not know that he was a researcher:

> While I was getting to know nursing assistants and residents and experiencing aspects of their daily routines, I would surreptitiously take notes on scraps of paper, in the bathroom or otherwise out of sight, jotting down what someone had said or done. (pp. 6–7)

Thorne (1993) was able to take notes openly:

> I went through the school days with a small spiral notebook in hand, jotting descriptions that I later expanded into field notes. When I was at the margins of a scene, I took notes on the spot. When I was more fully involved, sitting and

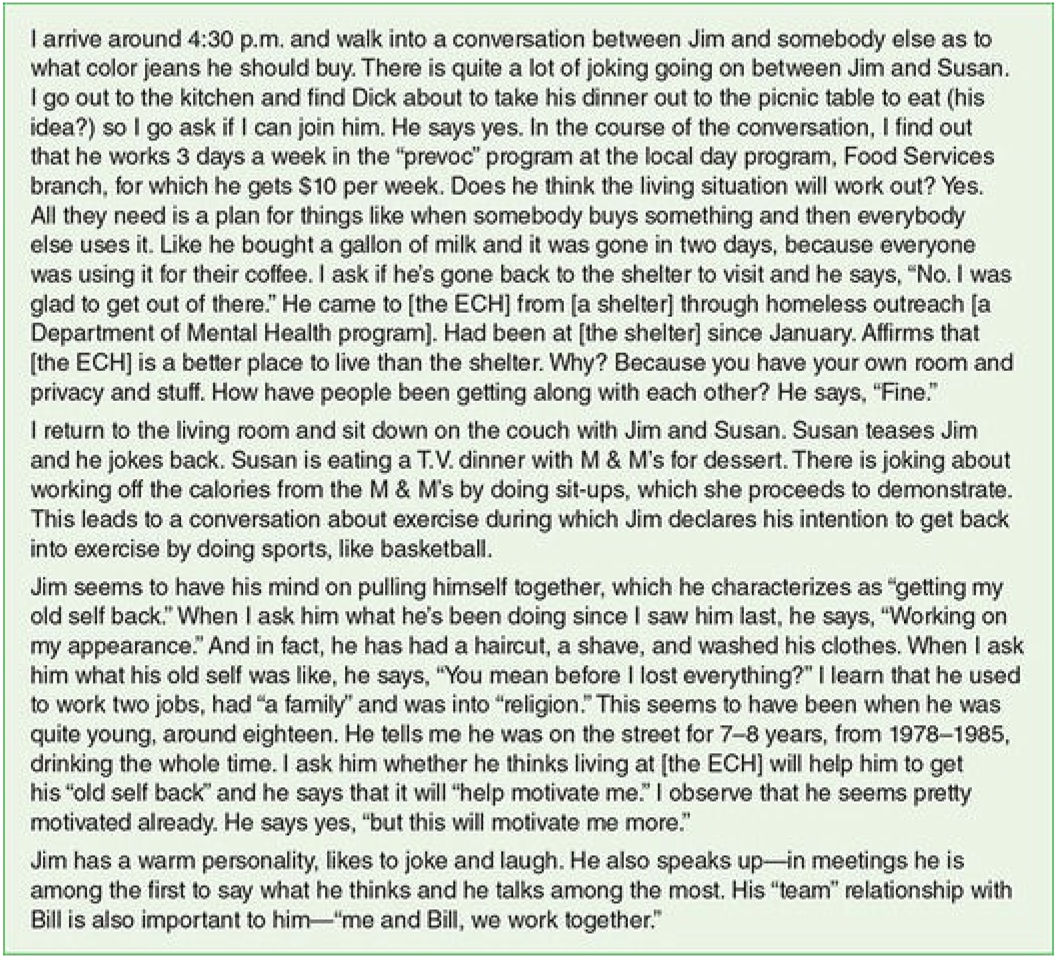
> talking with kids at a cafeteria table or playing a game of jump rope, I held observations in my memory and recorded them later. (p. 17)

Usually, writing up notes takes much longer—at least three times longer—than the observing. Field notes must be as complete, detailed, and true as possible to what was observed and heard. Direct quotes should be distinguished clearly from paraphrased quotes, and both should be set off from the researcher's observation and reflections. Pauses and interruptions should be indicated. The surrounding context should receive as much attention as possible, and a map of the setting should always be included with indications of where the individuals were at different times. The following excerpt from field notes collected by the DRC team members in the government's JFO show how notes can preserve a picture of the context:

> In the course of several weeks the [JFO] building has been wired to accommodate the increased electrical needs and the computer needs of the personnel. People were sleeping in bunk beds on site, in closets, and in corners of any room. The operation runs 24–7. Maps are hung on almost every wall with every type of imaginable data, from flooded areas to surge areas; total population and population density; number of housing, buildings, and people impacted by Katrina. . . . There is a logistics supply store that is full of materials and supplies, with a sign reminding people to take only what they need and this was a "no looting zone." The DRC team also observed flyers focusing on "stress management," as well as "how to cope with over-stressed workers." (Rodríguez et al. 2006:96)

Careful note taking yields a big payoff. On page after page, field notes will suggest new concepts, causal connections, and theoretical propositions. Social processes and settings can be described in rich detail, with ample illustrations. Exhibit 10.7, for example, contains field notes recorded by Norma Ware, an anthropologist studying living arrangements for homeless mentally ill persons in the Boston housing study for which I was a coinvestigator (Schutt 2011b). The notes contain observations of the setting, the questions the anthropologist asked, the answers she received, and her analytic thoughts about one of the residents. What can be learned from just this one page of field notes? The mood of the house at this time is evident, with joking, casual conversation, and close friendships. "Dick" remarks on problems with household financial management, and, at the same time, we learn a bit about his own activities and personality (a regular worker who appears to like systematic plans). We see how a few questions and a private conversation elicit information about the transition from the shelter to the house, as well as about household operations. The field notes also provide the foundation for a more complete picture of one resident, describing "Jim's" relationships with others, his personal history, his interests and personality, and his orientation to the future. We can also see analytic concepts emerge in the notes, such as the concept of pulling himself together, and of some house members working as a team. You can imagine how researchers can go on to develop a theoretical framework for understanding the setting and a set of concepts and questions to inform subsequent observations.

Exhibit 10.7 Field Notes From an Evolving Consumer Household (ECH)



I arrive around 4:30 p.m. and walk into a conversation between Jim and somebody else as to what color jeans he should buy. There is quite a lot of joking going on between Jim and Susan. I go out to the kitchen and find Dick about to take his dinner out to the picnic table to eat (his idea?) so I go ask if I can join him. He says yes. In the course of the conversation, I find out that he works 3 days a week in the "prevoc" program at the local day program, Food Services branch, for which he gets $10 per week. Does he think the living situation will work out? Yes. All they need is a plan for things like when somebody buys something and then everybody else uses it. Like he bought a gallon of milk and it was gone in two days, because everyone was using it for their coffee. I ask if he's gone back to the shelter to visit and he says, "No. I was glad to get out of there." He came to [the ECH] from [a shelter] through homeless outreach [a Department of Mental Health program]. Had been at [the shelter] since January. Affirms that [the ECH] is a better place to live than the shelter. Why? Because you have your own room and privacy and stuff. How have people been getting along with each other? He says, "Fine."

I return to the living room and sit down on the couch with Jim and Susan. Susan teases Jim and he jokes back. Susan is eating a T.V. dinner with M & M's for dessert. There is joking about working off the calories from the M & M's by doing sit-ups, which she proceeds to demonstrate. This leads to a conversation about exercise during which Jim declares his intention to get back into exercise by doing sports, like basketball.

Jim seems to have his mind on pulling himself together, which he characterizes as "getting my old self back." When I ask him what he's been doing since I saw him last, he says, "Working on my appearance." And in fact, he has had a haircut, a shave, and washed his clothes. When I ask him what his old self was like, he says, "You mean before I lost everything?" I learn that he used to work two jobs, had "a family" and was into "religion." This seems to have been when he was quite young, around eighteen. He tells me he was on the street for 7–8 years, from 1978–1985, drinking the whole time. I ask him whether he thinks living at [the ECH] will help him to get his "old self back" and he says that it will "help motivate me." I observe that he seems pretty motivated already. He says yes, "but this will motivate me more."

Jim has a warm personality, likes to joke and laugh. He also speaks up—in meetings he is among the first to say what he thinks and he talks among the most. His "team" relationship with Bill is also important to him—"me and Bill, we work together."

Source: Field notes from an ECH made available by Norma Ware, unpublished ethnographic notes, 1991.

Complete field notes must provide even more than a record of what was observed or heard. Notes also should include descriptions of the methodology: where researchers were standing or sitting while they observed, how they chose people for conversation or observation, what counts of people or events they made and why. Sprinkled throughout the notes also should be a record of the researchers' feelings and thoughts while observing: when they were disgusted by some statement or act, when they felt threatened or intimidated, why their attention shifted from one group to another, and what ethical concerns arose. Notes such as these provide a foundation for later review of the likelihood of bias or of inattention to some salient features of the situation.

Notes may, in some situations, be supplemented by still pictures, videotapes, and printed material circulated or posted in the research setting. Such visual material can bring an entirely different qualitative dimension into the analysis and call attention to some features of the social situation and actors within it that were missed in the notes (Grady 1996). Commentary on this material can be integrated with the written notes (Bogdewic 1999:67–68).

Managing the Personal Dimensions

Our overview of participant observation would not be complete without considering its personal dimensions. Because field researchers become a part of the social situation they are studying, they cannot help but be affected on a personal, emotional level. At the same time, those being studied react to researchers not just as researchers but as personal acquaintances —often as friends, sometimes as personal rivals. Managing and learning from this personal side of field research is an important part of any project.

The impact of personal issues varies with the depth of researchers' involvement in the setting. The more involved researchers are in the multiple aspects of the ongoing social situation, the more important personal issues become and the greater the risk of "going native." Even when researchers acknowledge their role, "increased contact brings sympathy, and sympathy in its turn dulls the edge of criticism" (Fenno 1978:277). Fenno minimized this problem by returning frequently to the university and by avoiding involvement in the personal lives of the congressional representatives he was studying. To study the social life of "corner boys," however, Whyte could not stay so disengaged. He moved into an apartment with a Cornerville family and lived for about 4 years in the community he was investigating:

> The researcher, like his informants, is a social animal. He has a role to play, and he has his own personality needs that must be met in some degree if he is to function successfully. Where the researcher operates out of a university, just going into the field for a few hours at a time, he can keep his personal social life separate from field activity. His problem of role is not quite so complicated. If, on the other hand, the researcher is living for an extended period in the community he is studying, his personal life is inextricably mixed with his research. (Whyte 1955:279)

Gill (2004) tried not to get too inured to crime as she studied communities in the Dominican Republic: "After several weeks my initial fear faded and I found it necessary to make periodic 'reality checks' with supervisors and colleagues abroad, which remind a researcher desensitized to violence and crime not to get too comfortable" (p. 3).

Thorne (1993) wondered whether "my moments of remembering, the times when I felt like a ten-year-old girl, [were] a source of distortion or insight?" She concluded that they were both: "Memory, like observing, is a way of knowing and can be a rich resource," but "when my own responses . . . were driven by emotions like envy or aversion, they clearly obscured my ability to grasp the full social situation" (p. 26). Deborah Ceglowski (2002) found that

> The feelings well up in my throat when Brian [a child in the Head Start program she studied] asks me to hold his hand. It's the gut reaction to hearing Ruth [a staff member] tell about Brian's expression when they pull into his yard and his mother isn't there. It is the caring connection of sitting next to Steven [another child] and hearing him say, "I miss my mom." (p. 15)

There is no formula for successfully managing the personal dimension of a field research project. It is much more an art than a science and flows more from the researcher's own personality and natural approach to other people than from formal training. Sharing similarities such as age, race, or gender with those who are studied may help create mutual feelings of comfort, but such social similarities may mask more important differences in perspective resulting from education, social class, and having the role of researcher (Doucet and Mauthner 2008:334). Furthermore, novice field researchers often neglect to consider how they will manage personal relationships when they plan and carry out their projects. Then, suddenly, they find themselves doing something they don't believe they should just to stay in the good graces of research subjects or juggling the emotions resulting from conflict within the group. As Whyte (1955) noted,

> The field worker cannot afford to think only of learning to live with others in the field. He has to continue living with himself. If the participant observer finds himself engaging in behavior that he has learned to think of as immoral, then he is likely to begin to wonder what sort of a person he is after all. Unless the field worker can carry with him a reasonably consistent picture of himself, he is likely to run into difficulties. (p. 317)

If you plan a field research project, follow these guidelines:

- Take the time to consider how you want to relate to your potential subjects as people.
- Speculate about what personal problems might arise and how you will respond to them.
- Keep in touch with other researchers and personal friends outside the research setting.
- Maintain standards of conduct that make you comfortable as a person and that respect the integrity of your subjects. (Whyte 1955:300–317)

When you evaluate participant observers' reports, pay attention to how they defined their role in the setting and dealt with personal problems. Don't place too much confidence in such research unless the report provides this information. The primary strengths of participant observation—learning about the social world from the participants' perspectives, as they experience it, and minimizing the distortion of these perspectives by the methods used to measure them—should not blind us to its primary weaknesses—the lack of consistency in the data collected, particularly when different observers are used, and the many opportunities for direct influence of the researchers' perspective on what is observed. Whenever we consider using the method of participant observation, we also must realize that the need to focus so much attention on each setting studied will severely restrict the possible number of settings or people we can study.

Intensive Interviewing

Intensive or depth interviewing is a qualitative method of finding out about people's experiences, thoughts, and feelings. Although intensive interviewing can be an important element in a participant observation study, it is often used by itself (Wolcott 1995:102–105). It shares with other qualitative research methods a commitment to learning about people in depth and on their own terms, and in the context of their situations.

Unlike the more structured interviewing that may be used in survey research (discussed in Chapter 8), intensive or depth interviewing relies on open-ended questions. Rather than asking standard questions in a fixed order, intensive interviewers may allow the specific content and order of questions to vary from one interviewee to another. Rather than presenting fixed responses that presume awareness of the range of answers that respondents might give, intensive interviewers expect respondents to answer questions in their own words.

What distinguishes intensive interviewing from less structured forms of questioning is consistency and thoroughness. The goal is to develop a comprehensive picture of the interviewee's background, attitudes, and actions, in his or her own terms, that is, to "listen to people as they describe how they understand the worlds in which they live and work" (Rubin and Rubin 1995:3). For example, Moore and his colleagues (2004) sought through intensive interviewing of key community leaders "to elicit a more general discussion on countywide events during the [Hurricane Floyd] flooding" (p. 209). The DRC researchers paid special attention to "firsthand personal accounts by individuals speaking about their own behavior" (Rodríguez et al. 2006:86).

Intensive interview studies do not reveal as directly as does participant observation the social context in which action is taken and opinions are formed. Nonetheless, intensive depth interviewers seek to account for context. Jack Douglas (1985) made the point succinctly in *Creative Interviewing:*

> *Creative interviewing is purposefully situated interviewing.* Rather than denying or failing to see the situation of the interview as a determinant of what goes in the questioning and answering processes, creative interviewing embraces the immediate, concrete situation; tries to understand how it is affecting what is communicated; and, by understanding these effects, changes the interviewer's communication processes to increase the discovery of the truth about human beings. (p. 22)

So, similar to participant observation studies, intensive interviewing engages researchers more actively with subjects than standard survey research does. The researchers must listen to lengthy explanations, ask follow-up questions tailored to the preceding answers, and seek to learn about interrelated belief systems or personal approaches to things rather than measure a limited set of variables. As a result, intensive interviews are often much longer than standardized interviews, sometimes as long as 15 hours, conducted in several different sessions. The intensive interview becomes more like a conversation between partners than an interview between a researcher and a subject (Kaufman 1986:22–23). Some call it “a conversation with a purpose” (Rossman and Rallis 1998:126).

Intensive interviewers actively try to probe understandings and engage interviewees in a dialogue about what they mean by their comments. To prepare for this active interviewing, the interviewer should learn in advance about the setting to be studied. Preliminary discussion with key informants, inspection of written documents, and even a review of your own feelings about the setting can all help (Miller and Crabtree 1999c:94–96). Robert Bellah and colleagues (1985) elaborate on this aspect of intensive interviewing in a methodological appendix to their national best seller about U.S. individualism, *Habits of the Heart:*

> We did not, as in some scientific version of “Candid Camera,” seek to capture their beliefs and actions without our subjects being aware of us, rather, we sought to bring our preconceptions and questions into the conversation and to understand the answers we were receiving not only in terms of the language but also so far as we could discover, in the lives of those we were talking with. Though we did not seek to impose our ideas on those with whom we talked . . . , we did attempt to uncover assumptions, to make explicit what the person we were talking to might rather have left implicit. The interview as we employed it was active, Socratic. (p. 304)

The intensive interview follows a preplanned outline of topics. It may begin with a few simple questions that gather background information while building rapport. These are often followed by a few general grand tour questions that are meant to elicit lengthy narratives (Miller and Crabtree 1999c:96–99). Some projects may use relatively structured interviews, particularly when the focus is on developing knowledge about prior events or some narrowly defined topic. But more exploratory projects, particularly those aiming to learn about interviewees’ interpretations of the world, may let each interview flow in a unique direction in response to the interviewee’s experiences and interests (Kvale 1996:3–5; Rubin and Rubin 1995:6; Wolcott 1995:113–114). In either case, qualitative interviewers must adapt nimbly throughout the interview, paying attention to nonverbal cues, expressions with symbolic value, and the ebb and flow of the interviewee’s feelings and interests. “You have to be free to follow your data where they lead” (Rubin and Rubin 1995:64).

Random selection is rarely used to select respondents for intensive interviews, but the selection method still must be considered carefully. If interviewees are selected in a haphazard manner, as by speaking just to those who happen to be available at the time when the researcher is on site, the interviews are likely to be of less value than when a more purposive selection strategy is used. Researchers should try to select interviewees who are knowledgeable about the subject of the interview, who are open to talking, and who represent the range of perspectives (Rubin and Rubin 1995:65–92). Selection of new interviewees should continue, if possible, at least until the saturation point is reached, the point when new interviews seem to yield little additional information (see Exhibit 10.8). As new issues are uncovered, additional interviewees may be selected to represent different opinions about these issues.

Grand tour question: A broad question at the start of an interview that seeks to engage the respondent in the topic of interest.

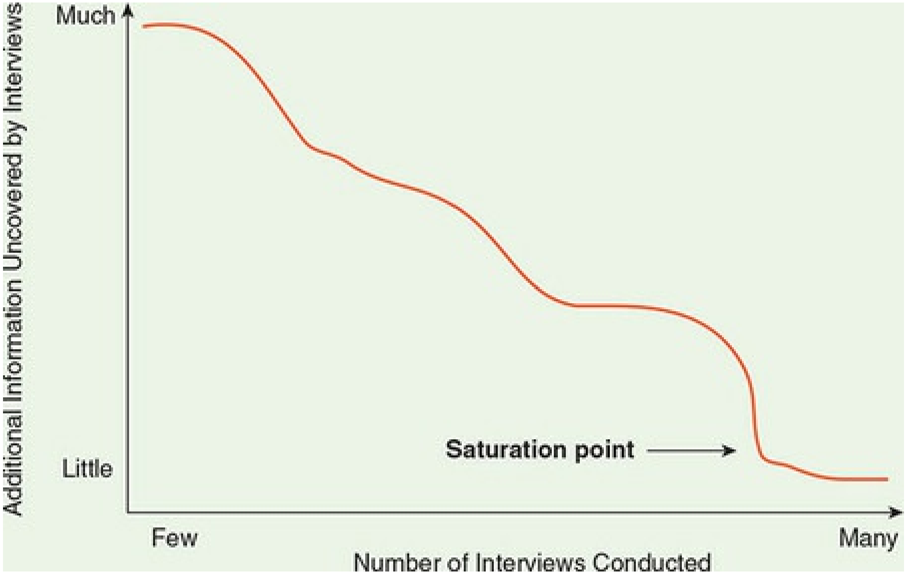
Saturation point: The point at which subject selection is ended in intensive interviewing, when new interviews seem to yield little additional information.

Establishing and Maintaining a Partnership

Because intensive interviewing does not engage researchers as participants in subjects' daily affairs, the problems of entering the field are much reduced. However, the social processes and logistics of arranging long periods for personal interviews can still be pretty complicated. Nordanger (2007:177) made a social visit to his Tigrayan interviewees before the interviews to establish trust and provide information about the interview. It also is important to establish rapport with subjects by considering in advance how they will react to the interview arrangements and by developing an approach that does not violate their standards for social behavior. Interviewees should be treated with respect, as knowledgeable partners whose time is valued (in other words, avoid coming late for appointments). A commitment to confidentiality should be stated and honored (Rubin and Rubin 1995).

But the intensive interviewer's relationship with the interviewee is not an equal partnership because the researcher seeks to gain certain types of information and strategizes throughout to maintain an appropriate relationship (Kvale 1996:6). In the first few minutes of the interview, the goal is to show interest in the interviewee and to explain clearly the purpose of the interview (p. 128). During the interview, the interviewer should maintain an appropriate distance from the interviewee, one that doesn't violate cultural norms; the interviewer should maintain eye contact—although this may not be appropriate in cultures in which eye contact is avoided—and not engage in distracting behavior. An appropriate pace is also important; pause to allow the interviewee to reflect, elaborate, and generally not feel rushed (Gordon 1992). When an interview covers emotional or otherwise stressful topics, the interviewer should give the interviewee an opportunity to unwind at the interview's end (Rubin and Rubin 1995:138).

Exhibit 10.8 The Saturation Point in Intensive Interviewing



More generally, intensive interviewers must be sensitive to the broader social context of their interaction with the interviewee and to the implications of their relationship in the way they ask questions and interpret answers. Tom Wengraf (2001) cautions new intensive interviewers to consider their unconscious orientations to others based on prior experience:

> Your [prior] experience of being interviewed may lead you to behave and "come across" in your interviewing . . . like a policeman, or a parent, a teacher or academic, or any "authority" by whom you have been interviewed and from whom you learnt a way of handling stress and ambiguity. (p. 18)

Asking Questions and Recording Answers

Intensive interviewers must plan their main questions around an outline of the interview topic. The questions should generally be short and to the point. More details can then be elicited through nondirective probes (e.g., "Can you tell me more about that?" or "uh-huh," echoing the respondent's comment, or just maintaining a moment of silence). Follow-up questions can then be tailored to answers to the main questions.

Interviewers should strategize throughout an interview about how best to achieve their objectives while accounting for interviewees' answers. *Habits of the Heart* again provides a useful illustration (Bellah et al. 1985:304):

> [Coinvestigator Steven] Tipton, in interviewing Margaret Oldham [a pseudonym], tried to discover at what point she would take responsibility for another human being:

Q: So what are you responsible for?

A: I'm responsible for my acts and for what I do.

Q: Does that mean you're responsible for others, too?

A: No.

Q: Are you your sister's keeper?

A: No.

Q: Your brother's keeper?

A: No.

Q: Are you responsible for your husband?

A: I'm not. He makes his own decisions. He is his own person. He acts his own acts. I can agree with them, or I can disagree with them. If I ever find them nauseous enough, I have a responsibility to leave and not deal with it any more.

Q: What about children?

A: I . . . I would say I have a legal responsibility for them, but in a sense I think they in turn are responsible for their own acts.

Do you see how the interviewer actively encouraged the subject to explain what she meant by “responsibility”? This sort of active questioning undoubtedly did a better job of clarifying her concept of responsibility than a fixed set of questions would have.

The active questioning involved in intensive interviewing, without a fixed interview script, also means that the statements made by the interviewee can only be understood in the context of the interviewer’s questions. Wengraf’s (2001:28–30) advice to a novice interviewer provides an example of how the interviewer’s statements and questions can affect the interview process:

Interviewer: Thank you for giving up this time for me.

Interviewee: Well, I don’t see it as giving up the time, more as contributing . . .

Interviewer: Well, for giving me the time, contributing the time, thank you very much.

Wengraf: Stay silent, let him clarify whatever the point is he wishes to make.
[Later],

Interviewer: Ok, so you’re anonymous [repeating an earlier statement], so you can say what you like.

Wengraf: Don’t imply that he is slow on the uptake—it might be better to cut the whole sentence.
[Later],

Interviewee: I guess, it’s difficult . . . being the breadwinner. Myself as a father, er . . . I’m not sure.

Wengraf: He’s uncertain, hoping to be asked some more about “being a father himself today.”

Interviewer Perhaps you could tell me a little about your own father.
[Slightly desperately]:

Wengraf: Interviewer ignores implied request but moves eagerly on to her intended focus on him as a son.

You can see in this excerpt how, at every step, the interviewer is actively constructing with the interviewee the text of the interview that will be analyzed. Becoming a good intensive interviewer means learning how to “get out of the way” as much as possible in this process. Becoming an informed critic of intensive interview studies means, in part, learning to consider how the social interaction between the interviewer and interviewee may have shaped in subtle ways what the interviewee said and to look for some report on how this interaction was managed.

Audio recorders commonly are used to record intensive and focus group interviews. Most researchers who have recorded interviews (including me) feel that they do not inhibit most interviewees and, in fact, are routinely ignored. The occasional respondent is very concerned with his or her public image and may therefore speak “for the audio recorder,” but such individuals are unlikely to speak frankly in any research interview. In any case, constant note taking during an interview prevents adequate displays of interest and appreciation by the interviewer and hinders the degree of concentration that results in the best interviews.

Of course, there are exceptions to every rule. Fenno (1978) presents a compelling argument for avoiding the audio recorder when interviewing public figures who are concerned with their public image:

> My belief is that the only chance to get a nonroutine, nonreflexive interview [from many of the members of Congress] is to converse casually, pursuing targets of opportunity without the presence of a recording instrument other than myself. If [worse] comes to worst, they can always deny what they have said in person; on tape they leave themselves no room for escape. I believe they are not unaware of the difference. (p. 280)

Interviewing Online

Our social world now includes many connections initiated and maintained through e-mail and other forms of web-based communication, so it is only natural that interviewing has also moved online. Online interviewing can facilitate interviews with others who are separated by physical distance; it also is a means to conduct research with those who are known only through such online connections as a discussion group, an e-mail distribution list, or social media (James and Busher 2009:14; Salmons 2012). As with digital ethnography, however, it is important to keep in mind how the online interview experience differs from face-to-face interviews and to consider the larger social context in which it takes place.

Online interviews can be either *synchronous*—in which the interviewer and interviewee exchange messages as in online chatting or with text messages—or *asynchronous*—in which the interviewee can respond to the interviewer's questions whenever it is convenient, usually through e-mail, but perhaps through a blog, a wiki, or an online forum (Salmons 2012). Both styles of online interviewing have advantages and disadvantages (James and Busher 2009:13–16). Synchronous interviewing provides an experience more similar to an in-person interview, thus giving more of a sense of obtaining spontaneous reactions, but it requires careful attention to arrangements and is prone to interruptions. Asynchronous interviewing allows interviewees to provide more thoughtful and developed answers, but it may be difficult to maintain interest and engagement if the exchanges continue over many days. The online asynchronous interviewer should plan carefully how to build rapport as well as how to terminate the online relationship after the interview is concluded (King and Horrocks 2010:86–93). Adding video to the exchange can also increase engagement, whether through real-time videoconferencing or by sending video clips or podcasts (Salmons 2012:4–5).

Allison Deegan (2012) initially tried synchronous interviewing in her study of teenage girls who had participated in the WriteGirl mentoring program in Los Angeles. She had learned that most program alumnae used Facebook and so contacted them on Facebook and began to set up "chat" sessions. However, her first interviews turned out to be too slow and seemed jerky, so she instead began arranging asynchronous interviews by e-mail. Respondents could complete the interview questions and return the interview form when it was convenient for them. Deegan learned later that the pace of the chat sessions was slow in part because respondents were doing multiple other tasks while they were in the sessions. So although the asynchronous approach took longer, it allowed for a better interview with a more focused respondent.

Whether a synchronous or asynchronous approach is used, online interviewing can facilitate the research process by creating a written record of the entire interaction without the need for typed transcripts. The relative anonymity of online communications can also encourage interviewees to be more open and honest about their feelings than they would be if interviewed in person (James and Busher 2009:24–25). However, online interviewing lacks some of the most appealing elements of qualitative methods: The revealing subtleties of facial expression, intonation, and body language are lost unless video is also added, and the intimate rapport that a good intensive interviewer can develop in a face-to-face interview cannot be achieved. In addition, those who are being interviewed have much greater ability to present an identity that is completely removed from their in-person persona; for instance, basic characteristics such as age, gender, and physical location can be completely misrepresented.

But if people are creating personas online to connect with others, that too becomes an important part of the social world to investigate—even if these online personas differ from the people who create them. Second Life is a three-dimensional virtual world in which real people are represented by avatars they create or purchase. Such virtual environments are part of the social world that millions of users experience and so are starting to become objects of investigation by qualitative researchers. In such virtual worlds, avatars interact with others in communities of various types and can buy and sell clothes, equipment, and buildings, and do almost anything that people do in the real world.

Ann Randall (2012) wondered whether Second Life members would experience beneficial effects from being interviewed in Second Life through their avatars, in the same way that "real people" report benefits from participating in qualitative interviews. To explore this issue, Randall e-mailed an invitation to Second Life members who were included on lists for educators and educational researchers. Randall's avatar then arranged to interview the avatars of those who agreed to be interviewed by returning a consent form with their avatar's name (retaining the anonymity of the interview). She then arranged for interviews with an availability sample of nine experienced Second Life members. Three chose to be interviewed in their own Second Life homes while the other six came to virtual locations that Randall created for the purpose, such as a beach house with reclining chairs.

The interview began with a friendly introductory exchange like the following between Randall's avatar and the participant's avatar (Randall 2012:139):

Ann: Hi!!

Participant: hello

Ann: nice to meet you

Participant: likewise

Ann: feel free to have a seat

Participant: thanks

Ann: you can sit on the stump if you want. I've restricted access so we can speak privately. Oh and have some coffee if you'd like. Right click on the coffee pot.

The interviews themselves were semi-structured, with a set of basic questions but the opportunity to pursue issues with additional questions as they arose. Six months later, Randall scheduled interviews with the same nine participants to ask them about possible beneficial effects of the interview. She reports finding much evidence of such beneficial effects, just as if actual people had been interviewed in person.

Randall (2012:144–147) emphasized the importance of exploring the specific virtual world(s) in which interviews would be conducted and of maintaining respectful communications based on the knowledge that each avatar represents a person with feelings that may not be expressed openly but that could be affected by the reactions of others. She also found that it was important to learn the Second Life "culture" before conducting research. One respondent drove this point home by explaining why the interviewer's avatar needed to dress and act appropriately (Randall 2012:143):

> Here where ppl [people] can look just how they want [i.e., in Second Life] . . . is where a lot of educators lose credibility. Most . . . never worry about shape, skin, hair. . . . So—you walk like a duck, stand like a stick, look like a dork, and wonder why people think you're an idiot?

Focus Groups

Focus groups are groups of unrelated individuals that are formed by a researcher and then led into group discussion of a topic for 1 to 2 hours (Krueger and Casey 2009:2). The researcher asks specific questions and guides the discussion to ensure that group members address these questions, but the resulting information is qualitative and relatively unstructured. Focus groups do not involve representative samples; instead, a few individuals are recruited who have the time to participate, have some knowledge pertinent to the focus group topic, and share key characteristics with the target population. Focus group research projects usually involve several discussions involving similar participants.

Focus groups have their roots in the interviewing techniques developed in the 1930s by sociologists and psychologists who were dissatisfied with traditional surveys. Traditionally, in a questionnaire survey, subjects are directed to consider certain issues and particular response options in a predetermined order. The spontaneous exchange and development of ideas that characterize social life outside the survey situation is lost—and with it, some social scientists fear, the prospects for validity.

During World War II, the military used focus groups to investigate morale, and then the great American sociologist Robert K. Merton and two collaborators, Marjorie Fiske and Patricia Kendall, popularized them in *The Focused Interview* (1956). But marketing researchers were the first to adopt focus groups as a widespread methodology. Marketing researchers use focus groups to investigate likely popular reactions to possible advertising themes and techniques. Their success has prompted other social scientists to use focus groups to evaluate social programs and to assess social needs (Krueger and Casey 2009:3–4).

Focus groups are now used extensively in political campaigns, as a quick means of generating insight into voter preferences and reactions to possible candidate positions. For example, Democratic Michigan legislators used focus groups to determine why voters were turning away from them in 1985. Elizabeth Kolbert (1992) found that white, middle-class Democrats were shifting to the Republican Party because of their feelings about race:

> These Democratic defectors saw affirmative action as a direct threat to their own livelihoods, and they saw the black-majority city of Detroit as a sinkhole into which their tax dollars were disappearing. . . . The participants listen[ed] to a quotation from Robert Kennedy exhorting whites to honor their "special obligation" to blacks. Virtually every participant in the four groups—37 in all—reacted angrily. (p. 21)

Focus groups are used to collect qualitative data, using open-ended questions posed by the researcher (or group leader). Thus, a focused discussion mimics the natural process of forming and expressing opinions. The researcher, or group moderator, uses an interview guide, but the dynamics of group discussion often require changes in the order and manner in which different topics are addressed (Brown 1999:120). No formal procedure exists for determining the generalizability of focus group answers, but the careful researcher should conduct at least several focus groups on the same topic and check for consistency in the findings. Some focus group experts advise conducting enough focus groups to reach the point of saturation, when an additional focus group adds little new information to that which already has been generated (Brown 1999:118). When differences in attitudes between different types of people are a concern, separate focus groups may be conducted that include these different types, and then the analyst can compare comments between them (Krueger and Casey 2009:21).

Most focus groups involve 5–10 people, a number that facilitates discussion by all in attendance (Krueger and Casey 2009:6). Participants usually do not know one another, although some studies in organized settings may include friends or coworkers. Opinions differ on the value of using homogeneous versus heterogeneous participants. Homogeneous groups may be more convivial and willing to share feelings, but heterogeneous groups may stimulate more ideas (Brown 1999:115–117). In any case, it is important to avoid having some participants who have supervisory or other forms of authority over other participants (Krueger and Casey 2009:22). It is also good to avoid focus groups when dealing with emotionally charged issues, when sensitive information is needed and when confidentiality cannot be ensured, or when the goal is to reach consensus (Krueger and Casey 2009:20).

Focus group moderators must begin the discussion by generating interest in the topic, creating the expectation that all will participate, and making it clear that the researcher does not favor any particular perspective or participant (Smithson 2008:361). All questions should be clear, simple, and straightforward. The moderator should begin with easy-to-answer general factual questions and then, about one quarter to halfway through the allotted time, shift to key questions on specific issues. In some cases, discussion may be stimulated by asking participants to make a list of concerns or experiences, to rate predefined items, or to choose between alternatives. If the question flow is successful, the participants should experience the focus group as an engaging and friendly discussion and the moderator should spend more time after the introductory period listening and guiding the discussion than asking questions. Disagreements should be handled carefully so that no participants feel uncomfortable, and the discussion should be kept focused on the announced topic (Smithson 2008:361). The moderator may conclude the group discussion by asking participants for recommendations to policy makers or their further thoughts that they have not had a chance to express (Krueger and Casey 2009:36–48).

Keith Elder and his colleagues at the University of South Carolina and elsewhere (2007:S125) used focus groups to study the decisions by African Americans not to evacuate

New Orleans before Hurricane Katrina. Elder et al. conducted six focus groups with 53 evacuees who were living in hotels in Columbia, South Carolina, between October 3 and October 14, 2005. African American women conducted the focus groups after American Red Cross relief coordinators announced them at a weekly "town hall" meeting.

One of the themes identified in the focus groups was the confusion resulting from inconsistent messages about the storm's likely severity. "Participants reported confusion about what to do because of inappropriate timing of mandatory evacuation orders and confusing recommendations from different authorities" (Elder et al. 2007:S126):

> The mayor did not say it was a mandatory evacuation at first. One or two days before the hurricane hit, he said it was mandatory. It was too late then.
>
> They didn't give us no warning. . . . When they said leave, it was already too late.
>
> After [the] levees broke the mayor said mandatory evacuation, before then he was not saying mandatory evacuation.
>
> Governor said on TV, you didn't want to go, you didn't have to go, cause it was no threat to us, she said.

Focus group methods share with other field research techniques an emphasis on discovering unanticipated findings and exploring hidden meanings. They can be an indispensable aid for developing hypotheses and survey questions, for investigating the meaning of survey results, and for quickly assessing the range of opinion about an issue. The group discussion reveals the language participants used to discuss topics and think about their experiences (Smithson 2008:359). However, in part due to their use in marketing and campaign research, focus groups have been viewed with some suspicion as a research technique and as creating a tendency to "groupthink" that suppresses open expression of individual opinions (Puchta and Potter 2004:75–76). Careful planning of group moderation to encourage diverse opinions and even including trained "devil's advocates" to argue for contrary opinions can help to encourage expression of alternative viewpoints (MacDougall and Baum 1997). Because it is not possible to conduct focus groups with large, representative samples, it is always important to consider how recruitment procedures have shaped the generalizability of focus group findings. The issue of impact of interviewer style and questioning on intensive interview findings, which was discussed in the previous section, also must be considered when evaluating the results of focus groups.

Generalizability in Qualitative Research

Qualitative research often focuses on populations that are hard to locate or are very limited in size. In consequence, nonprobability sampling methods such as availability sampling and snowball sampling are often used (see Chapter 5). However, this does not mean that generalizability should be ignored in qualitative research (Gobo 2008:206). Janet Ward Schofield (2002) suggests two different ways of increasing generalizability in qualitative investigations:

> *Studying the Typical.* Choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience. (p. 181)
> *Performing Multisite Studies.* A finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites. . . . Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more . . . likely to be useful in understanding various other sites than one emerging from the study of several very similar sites. (p. 184)

Giampietro Gobo (2008:204–205) highlights another approach to improving generalizability in qualitative research. A case may be selected for in-depth study because it is atypical, or deviant. Investigating social processes in a situation that differs from the norm will improve understanding of how social processes work in typical situations: "the exception that proves the rule."

Some qualitative researchers question the value of generalizability, as most researchers understand it. The argument is that understanding the particulars of a situation in depth is an important object of inquiry in itself. In the words of sociologist Norman Denzin,

> The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. . . . Every instance of social interaction . . . represents a slice from the life world that is the proper subject matter for interpretive inquiry. (Denzin cited in Schofield 2002:173)

You will have to decide for yourself whether it makes sense to be concerned with generalizability in qualitative research.

Ethical Issues in Qualitative Research

When qualitative researchers engage actively in the social world—in other words, when they are doing their job—they encounter unique ethical challenges. When a participant observer becomes an accepted part of a community or other group, natural social relations and sentiments will develop over time despite initial disclosure of the researcher's role. When a qualitative interviewer succeeds in gaining rapport, interviewees are likely to feel that they are sharing information with a caring friend rather than with a researcher to whom they gave their informed consent. When a well-run focus group leaves participants feeling they are in a natural conversation with acquaintances, spontaneous comments are likely to be made without consideration of their being recorded for research purposes. There is, then, "a certain unavoidable deception" (Bosk and De Vries 2004)

> that comes from trying to have both researcher and informant forget that what is going on is not a normal, natural exchange but research—not just everyday life as it naturally occurs but work, a job, a project. (p. 253)

The natural, evolving quality of much interaction between qualitative researchers and participants can make it difficult in many situations to stipulate in advance the procedures that will be followed to ensure ethical practice. In the words of Charles Bosk and Raymond De Vries (2004),

> Few of us start with specific hypotheses that we will later test in any systematic way. . . . We cannot state our procedures any more formally than we will hang around here in this particular neighborhood and try to figure out what is going on among these people. (p. 253)

Just as in everyday social interaction, a qualitative researcher's engagement with others "in the field" can go in a direction that was not anticipated in the formal research plan. Or as Martin Levinson (2010:195) put it, "the very nature of much ethnographic work—its openness and uncertainty—is such that a cunning researcher can easily circumvent ethical fences that are nominally in place." As an example, Levinson (2010:196–197) recounts his effort to gain the consent of a potential interviewee in his ethnographic study of Romani gypsies in the United Kingdom:

Me [Levinson]: So, would you mind answering a few questions?
No response.

Me: Well, is there anything you'd like to say?

Smithy: Yes—f—k off.
Now generally I respect subjects' wishes. . . . But on this occasion, I felt bloody- minded. It had been a frustrating day: a long drive, nothing to show. . . . I persisted.

Me: What I'm hoping is that this research will be of some use in . . .

Smithy: I'm not f—king interested in what you say your work's about. I told you to f—k off.

Me: Look, I tell you what. I won't even ask you any questions. Just talk to me for five minutes—about whatever you like, then I'll f—k off.
It seemed to me that his expression softened. . . . Perhaps, too, I had passed some test of maleness.

Smithy: Buy me a drink, and I'll talk to you.

It turned out that the pub where Levinson took "Smithy" for a pint had a "No Travellers" sign [i.e., gypsies not allowed] on its door, so Levinson's interview also bought Smithy access to an otherwise forbidden place.

Although all qualitative researchers must deal with these ethical challenges of qualitative methods, the specific ethical issues in a particular project vary with the type of qualitative methods used and the specific circumstances in which a project is conducted. Each of the following issues should be considered.

Voluntary participation.

The first step in ensuring that subjects are participating in a study voluntarily is to provide clear information about the research and the opportunity to decline to participate (King and Horrocks 2010:99). However, even when such information is provided, the ongoing interaction between the researcher and the participant can blur the meaning of voluntary participation:

> A skilled researcher can establish rapport and convince subjects to reveal almost anything, including things the researcher may not want to be responsible for knowing. (Sieber and Tolich 2013:164)

This problem of diminishing attention to the voluntary nature of participation can be reduced when qualitative interviews are conducted in multiple sessions by making clear to participants at each session their opportunity to withdraw consent (King and Horrocks 2010:115).

Maintaining the standard of voluntary participation can present more challenges in participant observation studies, particularly those using an ethnographic approach that immerses the researcher in the social world of the regular participants. Few researchers or institutional review boards (IRBs) are willing to condone covert participation because it offers no way to ensure that participation by the subjects is voluntary. However, interpreting the standard of voluntary participation can be difficult even when the researcher's role is more open. Practically, much field research would be impossible if participant observers were required to request permission of everyone in a group or setting, no matter how minimal their involvement (Calvey 2017:139). Levinson (2010:197) also explains that access to participants is not "some immutable, fixed state" that is resolved through gaining informed consent at the start of a project:

> The implication being that, once negotiated, access ceases to be an issue. Experience soon taught me that . . . what seemed acceptable one day was evidently not the next. Factors here ranged from external events—death in the family, illness, . . . to the mood of respondents on a given day. . . . Sometimes, I suspected people had been told not to speak to me.

Some researchers recommend adherence to a process consent standard to adjust the ideal of informed consent to the reality of the evolving nature of social relationships and research plans in qualitative projects. Process consent allows participants to change their decision about participating at any point by requiring that the researcher check with participants at each stage of the project about their willingness to continue in the project (Sieber and Tolich 2013:139). For example, before the conclusion of his ethnographic study of sidewalk book vendors in New York City, Mitch Duneier (1999) rented a room in which he read to participants the excerpts from his planned book that pertained to them and let them decide whether to give consent at that point.

Information about the research that is provided to research participants must also be tailored to their capacities, interests, and needs if they are able to give truly informed consent (Hammersley and Traianou 2012:95). Jim Birckhead commented that he "never felt that" the fundamentalist Christian serpent-handlers he studied in the southern United States "fully comprehended what I was actually doing in their midst" (as quoted in Hammersley and Traianou 2012:97).

Digital ethnography and other online qualitative research strategies also present unique challenges about voluntary consent. Chat room participants seem to be hostile to being monitored by researchers when it is disclosed, but not so much if it is not (Hewson et al. 2016:111). Dean Dabney, Laura Dugan, Volkan Topalli, and Richard Hollinger (2006) used closed circuit TV records of shoppers in their study of bias in identification of shoplifters in a drug store.

Process consent: An interpretation of the ethical standard of voluntary consent that allows participants to change their decision about participating at any point by requiring that the researcher check with participants at each stage of the project about their willingness to continue in the project.

Subject well-being.

Every field researcher should consider carefully before beginning a project how to avoid harm to subjects. Direct harm to the reputations or feelings of particular individuals is a primary concern. Researchers can avoid such harm, in part, by maintaining the confidentiality of research subjects. Researchers also must try to avoid affecting adversely the course of events while engaged in a setting. Whyte (1955:335–337) found himself regretting having recommended that a particular politician be allowed to speak to a social club he was observing because the speech led to serious dissension in the club and strains between Whyte and some of the club members. Similar concerns were raised about Sudhir Venkatesh's (2008) doctoral research about a gang in a Chicago housing project. Venkatesh had gained the trust of housing residents after being given access to them by a local gang leader and the project's manager and so was able to ask residents about their sources of unreported income, such as through babysitting, prostitution, and car repair. Venkatesh then reviewed with the gang leader what he had learned, leading the gang leader to extort more money from residents who had sources of income about which he had been unaware and leading Venkatesh to be shunned by the residents who had shared their income secrets with him (Sieber and Tolich 2013:88–90). David Calvey (2017:139) witnessed fights between bouncers he was studying and did not take action in response. How would you have handled these issues?

The well-being of the group or community studied as a whole should also be considered in relation to publication or other forms of reporting findings. Some of the Cornerville men read Whyte's book and felt discomfited by it (others found it enlightening). Some police accused Van Maanen of damaging their reputation with his studies. Carolyn Ellis (1995:159–161) returned to a Chesapeake fishing community in the hope of conducting a follow-up to her earlier ethnographic study; she was surprised to learn that residents had read excerpts from her book about the community (provided by another researcher!) and some were very upset with how they had been portrayed.

> "I thought we was friends, you and me, just talkin.' I didn't think you would put it in no book."
>
> "But I told people down here I was writing a book," I reply feebly.
>
> "But I still thought we was just talkin.' And you said we're dirty and don't know how to dress." . . ."It's my life, not anybody else's business. Weren't yours neither." (pp. 79–80)

Although such consequences could follow from any research, even from any public discourse, they are a particular concern for qualitative researchers who become during their research an accepted part of a group or community. Ellis (1995:87–89) decided that she should have considered how she would have felt if others had written in such ways about her and should have been more sensitive to her role in the community. Some indigenous groups and disadvantaged communities now require that researchers seek approval for all their procedures and request approval before releasing findings, to prevent harm to their culture and interests (Lincoln 2009:160–161).

These problems are less likely in intensive interviewing and focus groups, but researchers using these methods should try to identify negative feelings both after interviews and after reports are released and then help distressed subjects cope with their feelings through debriefing or referrals for professional help. Online interviewing can create additional challenges for interviews in which respondents become inappropriately personal over time (King and Horrocks 2010:100–101). Researchers should spend time in local settings before the research plan is finalized to identify a range of community stakeholders with diverse perspectives who can then be consulted about research plans (Bledsoe and Hopson 2009:400).

Identity disclosure.

We already have considered the problems of identity disclosure, particularly in the case of covert participation. Current ethical standards require informed consent of research subjects, and most would argue that this standard cannot be met in any meaningful way if researchers do not disclose fully their identity. But how much disclosure about the study is necessary, and how hard should researchers try to make sure that their research purposes are understood? In field research on Codependents Anonymous, Leslie Irvine (1998) found that the emphasis on anonymity and expectations for group discussion made it difficult to disclose her identity. Less-educated subjects may not readily comprehend what a researcher is or be able to weigh the possible consequences of the research for themselves. The intimacy of the researcher–participant relationship in much qualitative research makes it difficult to inject reminders about the research into ongoing social interaction (Mabry 2008:221). Must researchers always inform everyone of their identity as researchers? Consider this: In her study of female gamblers, Jun Li (2008) felt that she could not ethnically maintain her covert status when older women gamblers made concerted efforts to dissuade her from getting addicted. However, when Li disclosed her identity as a researcher, they stopped talking to her:

> Once female gamblers were made known of my research role, they started to view me differently, treating me as a suspicious outsider who should not be entrusted because I did not share their experiences. (p. 107)

What about disclosing a change in the researchers' interests and foci while their study is in progress? Can a balance be struck between the disclosure of critical facts and a coherent research strategy?

Digital ethnographers also have to be concerned with identity disclosure by participants in online communities who can easily misrepresent their own identities. Men may masquerade as women and children as adults, or vice versa, and yet the research should not proceed unless an effort is made to obtain informed and voluntary consent. If they are concerned about participation by individuals who do not meet criteria for a study, including minimal age requirements, digital ethnographers can at least attempt to identify those whose identity is not credible through inconsistencies in their postings (Kozinets 2010:151–154). Internet-based research can also violate the principles of voluntary participation and identity disclosure when researchers participate in discussions and record and analyze text but do not identify themselves as researchers (Jesnadum 2000). Digital ethnographers minimize the risks caused by these uncertainties by making their own identities known, stating clearly their expectations for participation, and providing an explicit informed consent letter that is available as discussion participants come and go (Denzin and Lincoln 2008:274). Janet Salmons (2016:11) recommends communicating research plans to the website host and posting an introduction to the study written in plain language.

Confidentiality.

Field researchers normally use fictitious names for the characters in their reports, but doing so does not always guarantee confidentiality to their research subjects. Individuals in the setting studied may be able to identify those whose actions are described and thus may become privy to some knowledge about their colleagues or neighbors that had formerly been kept from them. Therefore, researchers should make every effort to expunge possible identifying material from published information and to alter unimportant aspects of a description when necessary to prevent identity disclosure. In any case, no field research project should begin if some participants clearly will suffer serious harm by being identified in project publications. On the other hand, the American Sociological Association does not require informed consent from persons in public places who are being observed as a condition of research, although local IRBs may differ (Sieber and Tolich 2013:137–138).

Focus groups create a particular challenge because the researcher cannot guarantee that participants will not disclose information that others would like to be treated as confidential. Cultures differ in their openness to dissent and their sensitivity to public discussion of personal issues, so cross-cultural and cross-national focus groups exacerbate the difficulty of establishing appropriate standards for confidentiality (Smithson 2008:365–366). This risk can be reduced at the start of the focus group by reviewing a list of "dos and don'ts" with participants. Nonetheless, focus group methodology simply should not be used for very personally sensitive topics (Smithson 2008:360–361).

Although a digital ethnographer should maintain the same privacy standards as other qualitative researchers do, the retention of information on the web makes it difficult to protect confidentiality with traditional mechanisms. Anyone reading a distinctive quote based on online text can use a search engine to try to locate the original text. Distinctive online names intended to protect anonymity still identify individuals who may be the target of hurtful commentary (Kozinets 2010:143–145). Some users may think of their postings to some online community site as private (Markham 2008:274). For these reasons, digital ethnographers who study sensitive issues should go to great lengths to disguise the identity of the community they have studied as well as of its participants who are quoted.

Appropriate boundaries.

Maintaining appropriate boundaries between the researcher and research participants is a uniquely important issue in qualitative research projects that creates challenges for identity disclosure, subject well-being, and voluntary participation. You probably are familiar with this issue in the context of guidelines for ethical professional practice that seek to prevent abuse resulting from the inherent power disparities in relations between clinicians and patients and teachers and students. This is a special issue in qualitative research that involves reducing the distinction between the researcher and the research subject (or research participant). Qualitative researchers may seek to build rapport with those they plan to interview by expressing an interest in their concerns and conveying empathy for their situation.

> With deeper rapport, interviewees become more likely to explore their more intimate experiences and emotions. Yet they also become more likely to discover and disclose experiences and feelings which, upon reflection, they would have preferred to keep private from others . . . , or not to acknowledge even to themselves. (Duncombe and Jessop 2002:112)

Are researchers just "faking friendship" for the purpose of the research, as Jean Duncombe and Julie Jessop (2002) posed the dilemma in a book chapter titled "'Doing Rapport' and the Ethics of 'Faking Friendship'"? The long-term relationships that can develop in participant observation studies can make it seem natural for researchers to offer tangible help to research participants, such as helping take a child to school or lending funds. These involvements can in turn make it difficult to avoid becoming an advocate for the research participants, rather than a sympathetic observer. Duneier (1999) explains how he grappled with this issue and the way that he resolved it in his study of impoverished sidewalk book vendors in New York City:

> Could I show my deep appreciation for their struggles and gain their appreciation for my purposes as a sociologist without paying for some

> simulacrum of it? How could I communicate my purposes as a researcher without dollar bills and small change in my hand? . . . I knew that my salary (while not very high) was quite high compared to the going rate on the sidewalk. . . . But with time I did learn to say no, and to communicate the anguish I felt in giving such an answer.
>
> The question of how to avoid intervening when one cannot or should not do so is different from the question of whether and how to help when one can and should. At times I was asked to do things as simple as telling what I knew about the law, serving as a reference for a person on the sidewalk as he or she dealt with a landlord or potential landlord, helping someone with rent when he was about to be evicted, and on one occasion finding and paying for a lawyer. In these situations, I did everything I could to be helpful, but I never gave advice, opinions, or help beyond what was asked for. (p. 355)

Erich Goode (2002) decided he should redraw a common boundary between researcher and participant so that he could justify having sex with participants of the advocacy organization for severely obese women he was studying. He explains his decision partly by noting the effect on his ability to empathize with the women:

> For me, the reality of the stigma of obesity became far more poignantly real after socializing in NAAFA for over three years. Even more so, dating and sleeping with fat women dramatized that reality to such an intense pitch that, in concrete situations, my compassion had become a virtual chemical reaction. In accompanying my companions on dates, to a restaurant, a movie, or a NAAFA convention, I suffered from what Erving Goffman refers to as "courtesy stigma" (1963, p. 30), that is, guilt by association with a stigmatized person. My partners and I experienced the stares, the smirks, the obscene sniggering, the derisive comments ("What—is the circus in town?"). (p. 532)

But Goode (2002) paid a high price for this level of engagement because he was never able to write up his findings for publication. He found himself unable to juggle research and romantic roles and concluded,

> A major reason why I found it impossible to complete the project and write the book was that I was conflicted. My written version of the reality of fat sex would have to have been a lot less negative than my observations would have allowed. I didn't want to discredit the organization, its membership, and its assertion of the dignity and worth of all humanity, the extremely fat included. My head is abuzz with the din of the contradictions I saw and felt. I feel relieved finally not to be

> forced to sort it all out. I suspect that I was too personally involved with NAAFA and its membership to have made much sense of it all. I finally grant that there may be such a thing as sensory and empirical overload. (p. 526)

The value of maintaining ethical professional boundaries is thus a two-way street.

Researcher safety.

Research "in the field" about disasters or simply unfamiliar neighborhoods or nations should not begin until any potential risks to researcher safety have been evaluated. As Virginia Dickson-Swift, Erica James, Sandra Kippen, and Pranee Liamputtong (2008) note, risks in qualitative research can be emotional as well as physical:

> Look after yourself—have someone who you can debrief with, in fact have two people who you can debrief with. . . . Think there is always a risk in this research. You have got to remember that sometimes we go into spaces in people's lives that others have not been and that has the potential to be risky, both physically if the environment is not a safe one, or emotionally if the research affects you in some way. (pp. 137–138)

Qualitative methods may provide the only opportunity to learn about organized crime in Russian ports (Belousov et al. 2007) or street crime in the Dominican Republic (Gill 2004), but they should not be used if the physical risks to the researchers are unacceptably high. Safety needs to be considered at the time of designing the research, not as an afterthought on arriving in the research site. As Gill (2004) learned, such advance planning can require more investigation than just reading the local newspapers: "Due to the community's marginality, most crimes, including murders, were never reported in newspapers, making it impossible to have known the insecurity of the field site ahead of time" (p. 2).

But being realistic about evaluating risk does not mean simply accepting misleading assumptions about unfamiliar situations or communities. Reports of a widespread breakdown in law and order in New Orleans were broadcast repeatedly after Hurricane Katrina, but the DRC researchers found that most nontraditional behavior in that period was actually "prosocial," rather than antisocial (Rodríguez et al. 2006):

> One group named itself the "Robin Hood Looters." The core of this group consisted of eleven friends who, after getting their own families out of the area, decided to remain at some high ground and, after the floodwaters rose, commandeered boats and started to rescue their neighbors. . . . For about two

> weeks they kept searching in the area. . . . They foraged for food and water from abandoned homes, and hence their group name. Among the important norms that developed were that they were going to retrieve only survivors and not bodies and that group members would not carry weapons. The group also developed informal understandings with the police and the National Guard. (p. 91)

These ethical issues cannot be evaluated independently. The final decision to proceed must be made after weighing the relative benefits and risks to participants and considering how throughout the project voluntary participation can reasonably be ensured, identity disclosed to newcomers, researcher-induced harm largely prevented, confidentiality maintained, and risks and benefits clearly and honestly evaluated. Few qualitative research projects will be barred by consideration of these ethical issues, but almost all projects require careful attention to them. The more important concern for researchers is to identify the ethically troublesome aspects of their proposed research and resolve them before the project begins and to act on new ethical issues as they come up during the project.

Conclusions

Qualitative research allows the careful investigator to obtain a richer and more intimate view of the social world than is possible with more structured methods. It is not hard to understand why so many qualitative studies have become classics in the social science literature. And the emphases in qualitative research on inductive reasoning and incremental understanding help stimulate and inform other research approaches. Exploratory research to chart the dimensions of previously unstudied social settings and intensive investigations of the subjective meanings that motivate individual action are particularly well served by the techniques of participant observation, intensive interviewing, and focus groups.

The very characteristics that make qualitative research techniques so appealing restrict their use to a limited set of research problems. It is not possible to draw representative samples for study using participant observation, and, for this reason, the generalizability of any particular field study's results cannot really be known. Only the accumulation of findings from numerous qualitative studies permits confident generalization, but here again, the time and effort required to collect and analyze the data make it unlikely that many field research studies will be replicated.

Even if qualitative researchers made more of an effort to replicate key studies, their notion of developing and grounding explanations inductively in the observations made in a particular setting would hamper comparison of findings. Measurement reliability is thereby hindered, as are systematic tests for the validity of key indicators and formal tests for causal connections.

In the final analysis, qualitative research involves a mode of thinking and investigating different from that used in experimental and survey research. Qualitative research is inductive and idiographic, whereas experiments and surveys tend to be conducted in a deductive, quantitative, and nomothetic framework. Both approaches can help social scientists learn about the social world; the proficient researcher must be ready to use either. Qualitative data are often supplemented with counts of characteristics or activities. And as you have already seen, quantitative data are often enriched with written comments and observations, and focus groups have become a common tool of survey researchers seeking to develop their questionnaires. Thus, the distinction between qualitative and quantitative research techniques is not always clear-cut, and combining methods is often a good idea. I'll return to this in Chapter 12, on mixed methods.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Adaptive research design	371
Case study	372
Complete (or covert) participant	377
Complete (or overt) observer	377
Constructivism	370
Covert observer	377
Digital ethnography	375
Ethnography	373
Experience sampling method (ESM)	386
Field notes	387
Field research	367
Field researcher	378
Focus groups	376
Gatekeeper	382
Grand tour question	392
Hermeneutic circle	370
Intensive (in-depth) interviewing	376
Jottings	387
Key informant	383
Participant observation	376
Participant observer	377
Process consent	402
Reactive effects	378
Reflexivity	368
Saturation point	392
Theoretical sampling	385
Thick description	373

Highlights

- Qualitative methods are most useful in exploring new issues, investigating hard-to-study groups, and determining the meaning people give to their lives and actions. In addition, most social research projects can be improved, in some respects, by taking advantage of qualitative techniques.
- Qualitative researchers tend to develop ideas inductively, try to understand the social context and sequential nature of attitudes and actions, and explore the subjective meanings that participants attach to events. These researchers rely primarily on participant observation, intensive interviewing, and, in recent years, focus groups.
- The constructivist paradigm guides some qualitative researchers, and emphasizes the importance of exploring and representing the ways in which different stakeholders in a social setting construct their beliefs. Constructivists interact with research subjects to develop a shared perspective on the issue being studied.
- Case studies use thick description and other qualitative techniques to provide a holistic picture of a

setting or group.

- Ethnographers attempt to understand the culture of a group.
- Participant observers may adopt one of several roles for a particular research project. Each role represents a different balance between observing and participating. Many field researchers prefer a moderate role, participating as well as observing in a group but acknowledging publicly the researcher role. Such a role avoids the ethical issues that covert participation poses while still allowing the insights into the social world derived from participating directly in it. The role that the participant observer chooses should be based on an evaluation of the problems that are likely to arise from reactive effects and the ethical dilemmas of covert participation.
- Field researchers must develop strategies for entering the field, developing and maintaining relations in the field, sampling, and recording and analyzing data. Selection of sites or other units to study may reflect an emphasis on typical cases, deviant cases, or critical cases that can provide more information than others. Sampling techniques commonly used within sites or in selecting interviewees in field research include theoretical sampling, purposive sampling, snowball sampling, quota sampling, and, in special circumstances, random selection with the experience sampling method.
- Digital ethnographers use ethnographic techniques to study online communities.
- Recording and analyzing notes is a crucial step in field research. Jottings are used as brief reminders about events in the field, and daily logs are useful to chronicle the researcher's activities. Detailed field notes should be recorded and analyzed daily. Analysis of the notes can guide refinement of methods used in the field and of the concepts, indicators, and models developed to explain what has been observed.
- Theoretical sampling or experience sampling methods can improve generalizability of qualitative research findings.
- Intensive interviews involve open-ended questions and follow-up probes, with specific question content and order varying from one interview to another. Intensive interviews can supplement participant observation data.
- Focus groups combine elements of participant observation and intensive interviewing. They can increase the validity of attitude measurement by revealing what people say when they present their opinions in a group context instead of in the artificial one-on-one interview setting.
- Five ethical issues that should be given particular attention in field research concern: (1) voluntary participation, (2) subject well-being, (3) identity disclosure, (4) confidentiality, and (5) researcher safety. "Process consent" procedures may be appropriate in ongoing field research projects. Qualitative research conducted online, with discussion groups or e-mail traffic, raises special concerns about voluntary participation and identity disclosure.
- Adding qualitative elements to structured survey projects and experimental designs can enrich understanding of social processes.

Discussion Questions

1. You read in this chapter the statement by Maurice Punch (1994) that "the crux of the matter is that some deception, passive or active, enables you to get at data not obtainable by other means" (p. 91). What aspects of the social world would be difficult for participant observers to study without being covert? Are there any situations that would require the use of covert observation to gain access? What might you do as a participant observer to lessen access problems while still acknowledging your role as a researcher?
2. Review the experiments and surveys described in previous chapters. Pick one, and propose a field research design that would focus on the same research question but with participant observation techniques in a local setting. Propose the role that you would play in the setting, along the participant observation continuum, and explain why you would favor this role. Describe the stages of your field research study, including your plans for entering the field, developing and maintaining relationships, sampling, and recording and analyzing data. Then, discuss what you would expect your study to add to the findings resulting from the study described in the book.
3. Intensive interviews are the core of many qualitative research designs. How do they differ from the structured survey procedures that you studied in Chapter 8? What are their advantages and disadvantages over standardized interviewing? How does intensive interviewing differ from the qualitative method of participant observation? What are the advantages and disadvantages of these two methods?
4. Research on disasters poses a number of methodological challenges. In what ways are qualitative methods suited to disaster research? What particular qualitative methods would you have emphasized if you had been able to design research in New Orleans in the immediate aftermath of Hurricane Katrina? What unique challenges would you have confronted because of the nature of the disaster?

Practice Exercises

1. Conduct a brief observational study in a public location on campus where students congregate. A cafeteria, a building lobby, or a lounge would be ideal. You can sit and observe, taking occasional notes unobtrusively, without violating any expectations of privacy. Observe for 30 minutes. Write up field notes, being sure to include a description of the setting and a commentary on your own behavior and your reactions to what you observed.
2. Develop an interview guide that focuses on a research question addressed in one of the studies in this book. Using this guide, conduct an intensive interview with one person who is involved with the topic in some way. Take only brief notes during the interview, and then write up as complete a record of the interview as soon as you can immediately afterward. Turn in an evaluation of your performance as an interviewer and note taker, together with your notes.
3. Devise a plan for using a focus group to explore and explain student perspectives on some current event. How would you recruit students for the group? What types of students would you try to include? How would you introduce the topic and the method to the group? What questions would you ask? What problems would you anticipate, such as discord between focus group members or digressions from the chosen topic? How would you respond to these problems?
4. Find the “Qualitative Research” lesson in the “Interactive Exercises” link on the study site. Answer the questions in this lesson to review the types of ethical issues that can arise in the course of participant observation research and review the corresponding article.
5. Review postings to an online discussion group. How could you study this group using digital ethnography? What challenges would you encounter?

Ethics Questions

1. Should covert observation ever be allowed in social science research? Do you believe that social scientists should simply avoid conducting research on groups or individuals who refuse to admit researchers into their lives? Some have argued that members of privileged groups do not need to be protected from covert research by social scientists—that this restriction should be applied only to disadvantaged groups and individuals. Do you agree? Why or why not? Do you feel that Alfred's (1976) covert participation observation in the Satanist group was unethical and should not be allowed? Why or why not?
2. Should any requirements be imposed on researchers who seek to study other cultures, to ensure that procedures are appropriate and interpretations are culturally sensitive? What practices would you suggest for cross-cultural researchers to ensure that ethical guidelines are followed? (Consider the wording of consent forms and the procedures for gaining voluntary cooperation.)

Web Exercises

1. Check your library's online holdings to see if it subscribes to the online version of the *Annual Review of Sociology*. If it does, go to that site and search for articles that use qualitative methods as the primary method of gathering data on any one of the following subjects: child development/socialization; gender/sex roles; aging/gerontology. Enter "Qualitative AND Methods" in the subject field to begin this search. Review at least five articles, and report on the specific method of field research used in each.
2. Go to the QualPage site at https://qualpage.com/ and check out the resources listed. What information is provided regarding qualitative research, what kinds of qualitative projects are being published, and what purposes are specific qualitative methods being used for?
3. You have been asked to do field research on the World Wide Web's impact on the socialization of children in today's world. The first part of the project involves your writing a compare and contrast report on the differences between how you and your generation were socialized as children and the way children today are being socialized. Collect your data by surfing the web "as if you were a kid." The web is your field, and you are the field researcher.

Video Interview Questions

Listen to the researcher interviews for Chapter 10 at edge.sagepub.com/schutt9e.

1. What type of research design did Andrea Leverentz use in her study? What were some of the advantages and disadvantages of this type of design that were mentioned in the interview?
2. What new questions and issues came up during Leverentz’s research, and how did these differ from the original research question or focus? What does this say about the inductive approach and the importance of, as Leverentz says, letting “the data speak to you?”
3. According to Lakshmi Srinivas, what are the benefits to ethnographic research?
4. What challenges of ethnographic research does Srinivas highlight?

SPSS Exercises

The cross-tabulations you examined in Chapter 8's SPSS exercises highlighted the strength of the association between attitudes related to race and support for capital punishment. In this chapter, you will explore related issues. As you carry out this analysis, consider what additional information you might be able to obtain about these relationships with qualitative interviews of some of the respondents.

1. Examine the association between race and support for capital punishment. From the menu, click:
   Analyze/Descriptive Statistics/Crosstabs
   In the Crosstabs window, set
   Rows: cappun
   Columns: race
   Cells: column percents
2. What is the association between race and support for capital punishment? How would you explain that association?
3. Now consider what might lead to variation in support for capital punishment between whites and blacks. Consider gun ownership (OWNGUN), religious beliefs (FUND), attitudes about race (RACOPEN), education (EDUCSMALL), and political party identification (PARTYID3).
4. Generate crosstabs for the association of support for capital punishment with each of these variables, separately for minorities and whites. Follow the same procedures you used in Step 1, substituting the variables mentioned in Step 3 for RACE in Step 1. However, you must repeat the crosstab request for blacks and whites. To do this, before you choose Analyze, select black respondents only. From the menu above the Data Editor window, select Data, then Select Cases. Then from the Select Cases window, select If condition is satisfied and create this expression:
   If . . . RACE=1
   After you have generated the crosstabs, go back and repeat the data selection procedures, ending with RACE=2. When finished with the exercises, be sure to go back to Select Cases and select All Cases.
5. Are the bases of support for capital punishment similar among minorities and whites? Discuss your findings.
6. Propose a focus group to explore these issues further. Identify the setting and sample for the study, and describe how you would carry out your focus group.

Developing a Research Proposal

Add a qualitative component to your proposed study. You can choose to do this with a participant observation project or intensive interviewing. Pick the method that seems most likely to help answer the research question for the overall survey project (see Exhibit 3.10, #13 to #17).

1. For a participant observation component, propose an observational plan that would complement the overall survey project. Present in your proposal the following information about your plan: (a) Choose a site and justify its selection in terms of its likely value for the research; (b) choose a role along the participation–observation continuum and justify your choice; (c) describe access procedures and note any likely problems; (d) discuss how you will develop and maintain relations in the site; (e) review any sampling issues; and (f) present an overview of the way in which you will analyze the data you collect.
2. For an intensive interview component, propose a focus for the intensive interviews that you believe will add the most to findings from the survey project. Present in your proposal the following information about your plan: (a) Present and justify a method for selecting individuals to interview; (b) write out three introductory biographical questions and five grand tour questions for your interview schedule; (c) list at least six different probes you may use; (d) present and justify at least

two follow-up questions for one of your grand tour questions; and (e) explain what you expect this intensive interview component to add to your overall survey project.

Chapter 11 Qualitative Data Analysis

- Research That Matters, Questions That Count
- Features of Qualitative Data Analysis
  - Qualitative Data Analysis as an Art
  - Qualitative Compared With Quantitative Data Analysis
- Techniques of Qualitative Data Analysis
  - Documentation
  - Organization, Categorization, and Condensation
  - Examination and Display of Relationships
  - Corroboration and Legitimization of Conclusions
  - Reflection on the Researcher's Role
- Alternatives in Qualitative Data Analysis
  - Grounded Theory
- Research in the News: How to Understand Solitary Confinement
  - Abductive Analysis
  - Case-Oriented Understanding
  - Conversation Analysis
  - Narrative Analysis
  - Ethnomethodology
  - Qualitative Comparative Analysis
  - Combining Qualitative Methods
- Visual Sociology
- Careers and Research
- Systematic Observation
- Participatory Action Research
- Computer-Assisted Qualitative Data Analysis
- Ethics in Qualitative Data Analysis
- Conclusions

I was at lunch standing in line and he [another male student] came up to my face and started saying stuff and then he pushed me. I said . . . I'm cool with you, I'm your friend and then he push me again and calling me names. I told him to stop pushing me and then he push me hard and said something about my mom. And then he hit me, and I hit him back. After he fell I started kicking him.

—Calvin Morrill et al. (2000:521)

Research That Matters, Questions That Count

The Sexual Experiences Survey (SES) is used on many college campuses to assess the severity of sexual victimization, but researchers have found that it does not differentiate well between situations of unwanted sexual contact and attempted rape. Jenny Rinehart and Elizabeth Yeater (2011:927) at the University of New Mexico designed a project to develop "a deeper qualitative understanding of the details of the event, as well as the context surrounding it."

As part of a larger study of dating experiences at a West Coast university, Rinehart and Yeater analyzed written narratives provided by 78 women who had indicated some experience with sexual victimization on the SES. The authors and an undergraduate research assistant read each of the narratives and identified eight different themes and contexts, such as "relationship with the perpetrator" and "woman's relationship with the perpetrator." Next, they developed specific codes to make distinctions within each of the themes and contexts, such as "friend," "boss," or "stranger" within the "relationship" theme.

Here is an incident in one narrative that Rinehart and Yeater (2011:934) coded as involving unwanted sexual contact with a friend:

> I went out on a date with a guy (he was 24) and we had a good time. He invited me into his apartment after to "hang out" for a little while longer. He tried pressuring me into kissing him at first, even though I didn't want to. Then he wrestled me (playfully to him, but annoyingly and unwanted to me). I repeatedly asked him to get off of me, and eventually he did. I kissed him once.

Their analysis of these narratives made it clear that incidents that received the same SES severity rating often differed considerably when the particulars were examined.

1. According to the authors, "Grouping disparate events into the same category for ease of analysis or description may actually hinder researchers from truly understanding sexual victimization" (p. 939). Do you think that structured surveys can be refined to make the important distinctions, or should researchers shift more often to qualitative interviews when investigating the social world?
2. Do you think it would be preferable to analyze the narratives as text, rather than coding them into categories? Why or why not?

In this chapter, you will learn the language and techniques of qualitative data analysis, as well as a bit about research on victimization. By the end of the chapter, you will understand the distinctive elements of qualitative data analysis and how it differs from quantitative data analysis. After you finish the chapter, you can test yourself by reading the 2011 *Violence Against Women* article by Rinehart and Yeater at the *Investigating the Social World* study site and by completing the related interactive exercises for Chapter 11 at edge.sagepub.com/schutt9e.

Rinehart, Jenny K. and Elizabeth A. Yeater. 2011. "A Qualitative Analysis of Sexual Victimization Narratives." *Violence Against Women* 17(7):925–943.

Unfortunately, this statement was not made by a soap opera actor but by a real student writing an in-class essay about conflicts in which he had participated. But then you already knew that such conflicts are common in many high schools, so perhaps it will be reassuring to know that this statement was elicited by a team of social scientists who were studying conflicts in high schools to better understand their origins and to inform prevention policies.

The first difference between qualitative and quantitative data analysis is that the data to be analyzed are text, rather than numbers, at least when the analysis first begins. Does it trouble you to learn that there are no variables and hypotheses in this qualitative analysis by

Morrill et al. (2000)? This, too, is another difference between the typical qualitative and quantitative approaches to analysis, although there are some exceptions.

In this chapter, I present the features that most qualitative data analyses share, and I will illustrate these features with research on youth conflict and on homelessness. You will quickly learn that there is no one way to analyze textual data. To quote Michael Quinn Patton (2002), "Qualitative analysis transforms data into findings. No formula exists for that transformation. Guidance, yes. But no recipe. Direction can and will be offered, but the final destination remains unique for each inquirer, known only when—and if—arrived at" (p. 432).

I will discuss some of the different types of qualitative data analysis before focusing on computer programs for qualitative data analysis; you will see that these increasingly popular programs are blurring the distinctions between quantitative and qualitative approaches to textual analysis.

Features of Qualitative Data Analysis

The distinctive features of qualitative data collection methods that you studied in Chapter 10 are also reflected in the methods used to analyze those data. The focus on text—on qualitative data rather than on numbers—is the most important feature of qualitative analysis. The "text" that qualitative researchers analyze is most often transcripts of interviews or notes from participant observation sessions, but *text* can also refer to pictures or other images that the researcher examines.

What can the qualitative data analyst learn from a text? Here qualitative analysts may have two different goals. Some qualitative researchers view analysis of a text as a way to understand what participants "really" thought, felt, or did in a given situation or at a point in time. The text becomes a way to get "behind the numbers" that are recorded in a quantitative analysis to see the richness of real social experience. Others have adopted a hermeneutic perspective on texts—that is, a perspective that views a text as an interpretation that can never be judged true or false. The text is only one possible interpretation among many (Patton 2002:114).

From a hermeneutic perspective, then, the meaning of a text is negotiated among a community of interpreters, and to the extent that some agreement is reached about meaning at a particular time and place, that meaning is based on consensual community validation. A hermeneutic researcher is thus constructing a "reality" with his or her interpretations of a text provided by the subjects of research; other researchers, with different backgrounds, could come to markedly different conclusions.

You can see in this discussion about text that qualitative and quantitative data analyses also differ in the priority given to the preexisting views of the researcher and to those of the subjects of the research. Qualitative data analysts seek to describe their textual data in ways that capture the setting or people who produced this text on their own terms rather than in terms of predefined measures and hypotheses. What this means is that qualitative data analysis tends to be inductive—the analyst identifies important categories in the data, as well as patterns and relationships, through a process of discovery. There are often no predefined measures or hypotheses. Anthropologists term this an emic focus, which means representing the setting in terms of the participants and their viewpoint—an "insider" perspective based on the researcher's immersion in the setting. This focus contrasts with the etic focus of most quantitative research, in which the researcher seeks to remove himself or herself from the setting and represents the participants in the researcher's terms and in relation to a research question based on the preexisting literature (Salmons 2016:102).

Emic focus: Representing a setting with the participants' terms and from their viewpoint.

Etic focus: Representing a setting with the researchers' terms and from their viewpoint.

Good qualitative data analyses also are distinguished by their focus on the interrelated aspects of the setting, group, or person under investigation—the case—rather than breaking the whole into separate parts. The whole is always understood to be greater than the sum of its parts, and so the social context of events, thoughts, and actions becomes essential for interpretation. Within this framework, it doesn't really make sense to focus on two variables out of an interacting set of influences and test the relationship between just those two.

Qualitative data analysis is an iterative and reflexive process that begins as data are being collected rather than after data collection has ceased (Stake 1995). Next to her field notes or interview transcripts, the qualitative analyst jots down ideas about the meaning of the text and how it might relate to other issues. This process of reading through the data and interpreting them continues throughout the project. The analyst adjusts the data collection process itself when it begins to appear that additional concepts need to be investigated or new relationships explored. This process is termed progressive focusing (Parlett and Hamilton 1976).

> We emphasize placing an interpreter in the field to observe the workings of the case, one who records objectively what is happening but simultaneously examines its meaning and redirects observation to refine or substantiate those meanings. Initial research questions may be modified or even replaced in mid-study by the case researcher. The aim is to thoroughly understand [the case]. If early questions are not working, if new issues become apparent, the design is changed. (Stake 1995:9)

Elijah Anderson (2003) describes the progressive focusing process in his memoir about his study of Jelly's Bar:

> Throughout the study, I also wrote conceptual memos to myself to help sort out my findings. Usually no more than a page long, they represented theoretical insights that emerged from my engagement with the data in my field notes. As I gained tenable hypotheses and propositions, I began to listen and observe selectively, focusing on those events that I thought might bring me alive to my research interests and concerns. This method of dealing with the information I was receiving amounted to a kind of a dialogue with the data, sifting out ideas, weighing new notions against the reality with which I was faced there on the streets and back at my desk. (pp. 235–236)

Carrying out this process successfully is more likely if the analyst reviews a few basic guidelines when he or she starts the process of analyzing qualitative data (Miller and Crabtree 1999b:142–143):

- Know yourself, your biases, and preconceptions.
- Know your question.
- Seek creative abundance. Consult others and keep looking for alternative interpretations.
- Be flexible.
- Exhaust the data. Try to account for all the data in the texts, then publicly acknowledge the unexplained and remember the next principle.
- Celebrate anomalies. They are the windows to insight.
- Get critical feedback. The solo analyst is a great danger to self and others.
- Be explicit. Share the details with yourself, your team members, and your audiences.

Progressive focusing: The process by which a qualitative analyst interacts with the data and gradually refines his or her focus.

Qualitative Data Analysis as an Art

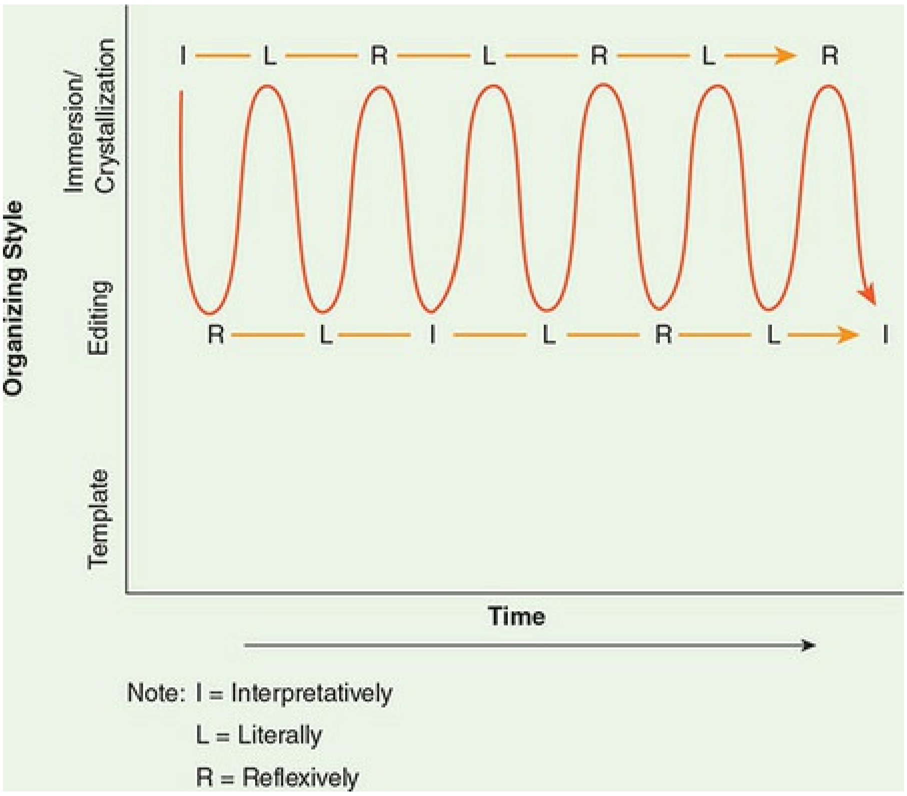
If you find yourself longing for the certainty of predefined measures and deductively derived hypotheses, you are beginning to understand the difference between setting out to analyze data quantitatively and planning to do so with a qualitative approach in mind. Or maybe you are now appreciating better the contrast between the positivist and constructivist research philosophies that I summarized in Chapter 1. When it comes right down to it, the process of qualitative data analysis is even described by some as involving as much "art" as science—as a "dance," in the words of William Miller and Benjamin Crabtree (1999b) (see Exhibit 11.1):

> Interpretation is a complex and dynamic craft, with as much creative artistry as technical exactitude, and it requires an abundance of patient plodding, fortitude, and discipline. There are many changing rhythms; multiple steps; moments of jubilation, revelation, and exasperation. . . . The dance of interpretation is a dance for two, but those two are often multiple and frequently changing, and there is always an audience, even if it is not always visible. Two dancers are the interpreters and the texts. (pp. 138–139)

Miller and Crabtree (1999b) identify three different modes of reading the text within the dance of qualitative data analysis:

1. When the researcher reads the text literally, she is focused on its literal content and form, so the text "leads" the dance.
2. When the researcher reads the text reflexively, she focuses on how her own orientation shapes her interpretations and focus. Now, the researcher leads the dance.
3. When the researcher reads the text interpretively, she tries to construct her own interpretation of what the text means.

Exhibit 11.1 Dance of Qualitative Analysis


Organizing Style
Immersion/ Crystallization
Editing
Template
Time
I — L — R — L — R — L → R
R — L — I — L — R — L → I
Note: I = Interpretatively
L = Literally
R = Reflexively

Source: Miller and Crabtree (1999b:139, Figure 7.1, based on Addison 1999). Reprinted with permission from SAGE Publications, Inc.

Sherry Turkle's (2011) book *Alone Together: Why We Expect More From Technology and Less From Each Other* provides many examples of this analytic dance, although of course in the published book we are no longer able to see that dance in terms of Turkle's original notes. She often describes what she observed in classrooms. Here's an example of such a *literal* focus, reflecting her experience in MIT's Media Lab at the start of the mobile computing revolution:

> In the summer of 1996, I met with seven young researchers at the MIT Media Lab who carried computers and radio transmitters in their backpacks and keyboards in their pockets. . . . They called themselves "cyborgs" and were always wirelessly connected to the Internet, always online, free from desks and cables. (Turkle 2011:151)

Such literal reports are interspersed with *interpretive* comments about the meaning of her observations:

> The cyborgs were a new kind of nomad, wandering in and out of the physical

> real. . . . The multiplicity of worlds before them set them apart; they could be with you, but they were always somewhere else as well. (Turkle 2011:152)

And several times in each chapter, Turkle (2011) makes *reflexive* comments on her own reactions:

> I don't like the feeling of always being on call. But now, with a daughter studying abroad who expects to reach me when she wants to reach me, I am grateful to be tethered to her through the Net. . . . Even these small things allow me to identify with the cyborgs' claims of an enhanced experience. Tethered to the Internet, the cyborgs felt like more than they could be without it. Like most people, I experience a pint-sized version of such pleasures. (p. 153)

In this artful way, the qualitative data analyst reports on her notes from observing or interviewing, interprets those notes, and considers how she reacts to the notes. These processes emerge from reading the notes and continue while she is editing the notes and deciding how to organize them, in an ongoing cycle.

Qualitative Compared With Quantitative Data Analysis

With this process in mind, let's review the many ways in which qualitative data analysis differs from quantitative analysis (Denzin and Lincoln 2000:8–10; Patton 2002:13–14). Each difference reflects the qualitative data analyst's orientation to in-depth, comprehensive understanding in which the analyst is an active participant compared with the quantitative data analyst's role as a dispassionate investigator of specific relations between discrete variables:

- A focus on meanings rather than on quantifiable phenomena
- Collection of many data on a few cases rather than few data on many cases
- Study in depth and detail, without predetermined categories or directions, rather than emphasis on analyses and categories determined in advance
- Conception of the researcher as an "instrument," rather than as the designer of objective instruments to measure particular variables
- Sensitivity to context rather than seeking universal generalizations
- Attention to the impact of the researcher's and others' values on the course of the analysis rather than presuming the possibility of value-free inquiry
- A goal of rich descriptions of the world rather than measurement of specific variables

As you learned in Chapter 8, the focus of qualitative data analysis on meaning and in-depth study also makes it a valuable supplement to analyses of quantitative data. Qualitative data like that collected in Renee Anspach's (1991) qualitative interviews in four community mental health systems can provide information about the quality of standardized case records and quantitative survey measures, as well as offer some insight into the meaning of particular fixed responses.

You'll also want to keep in mind features of qualitative data analysis that are shared with those of quantitative data analysis. Both qualitative and quantitative data analysis can involve making distinctions about textual data. Textual data can be transposed to quantitative data through a process of categorization and counting (see Chapter 14 on content analysis). Some qualitative analysts also share with quantitative researchers a positivist goal of describing better the world as it "really" is, although others have adopted a postmodern goal of trying to understand how different people see and make sense of the world, without believing that there is any "correct" description.

Techniques of Qualitative Data Analysis

Five different techniques are shared by most approaches to qualitative data analysis:

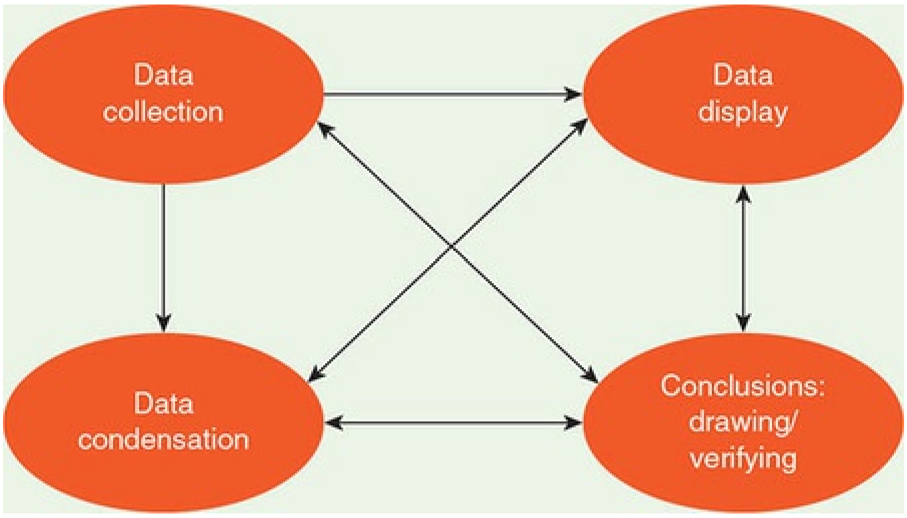
1. Documentation of the data and the process of data collection
2. Organization, categorization, and condensation of the data into concepts
3. Examination and display of relationships between concepts
4. Corroboration and legitimization of conclusions, by evaluating alternative explanations, disconfirming evidence, and searching for negative cases
5. Reflection on the researcher's role

Some researchers suggest different steps, or include additional steps, such as developing propositions that reflect the relationships found and making connections with extant theories (see Miles, Huberman, and Soldaña 2014:chap. 1). Exhibit 11.2 highlights the key techniques and emphasizes the reciprocal relations between them. In qualitative data analysis, condensation of data into concepts may lead to some conclusions and to a particular form of display of relationships between concepts, but the conclusions may then lead to changes in conceptualization and display, in an iterative process.

The analysis of qualitative research notes begins in the field, at the time of observation, interviewing, or both, as the researcher identifies problems and concepts that appear likely to help in understanding the situation. Simply reading the notes or transcripts is an important step in the analytic process. Researchers should make frequent notes in the margins to identify important statements and to propose ways of coding the data: "husband–wife conflict," perhaps, or "tension-reduction strategy."

An interim stage may consist of listing the concepts reflected in the notes and diagramming the relationships between concepts (Maxwell 2005:97–99). In large projects, weekly team meetings are an important part of this process. Susan Miller (1999) described this process in her study of neighborhood police officers (NPOs). Her research team met both to go over their field notes and to resolve points of confusion, as well as to dialogue with other skilled researchers who helped identify emerging concepts:

Exhibit 11.2 Components of Data Analysis: Interactive Model


Data collection
Data display
Data condensation
Conclusions: drawing/ verifying

Source: Miles, Huberman, and Saldaña (2014:chap. 1). Reprinted with permission from SAGE Publications, Inc.

> The fieldwork team met weekly to talk about situations that were unclear and to troubleshoot any problems. We also made use of peer-debriefing techniques. Here, multiple colleagues, who were familiar with qualitative data analysis but not involved in our research, participated in preliminary analysis of our findings. (p. 233)

This process continues throughout the project and should assist in refining concepts during the report-writing phase, long after data collection has ceased. Let's examine each of the stages of qualitative research in more detail.

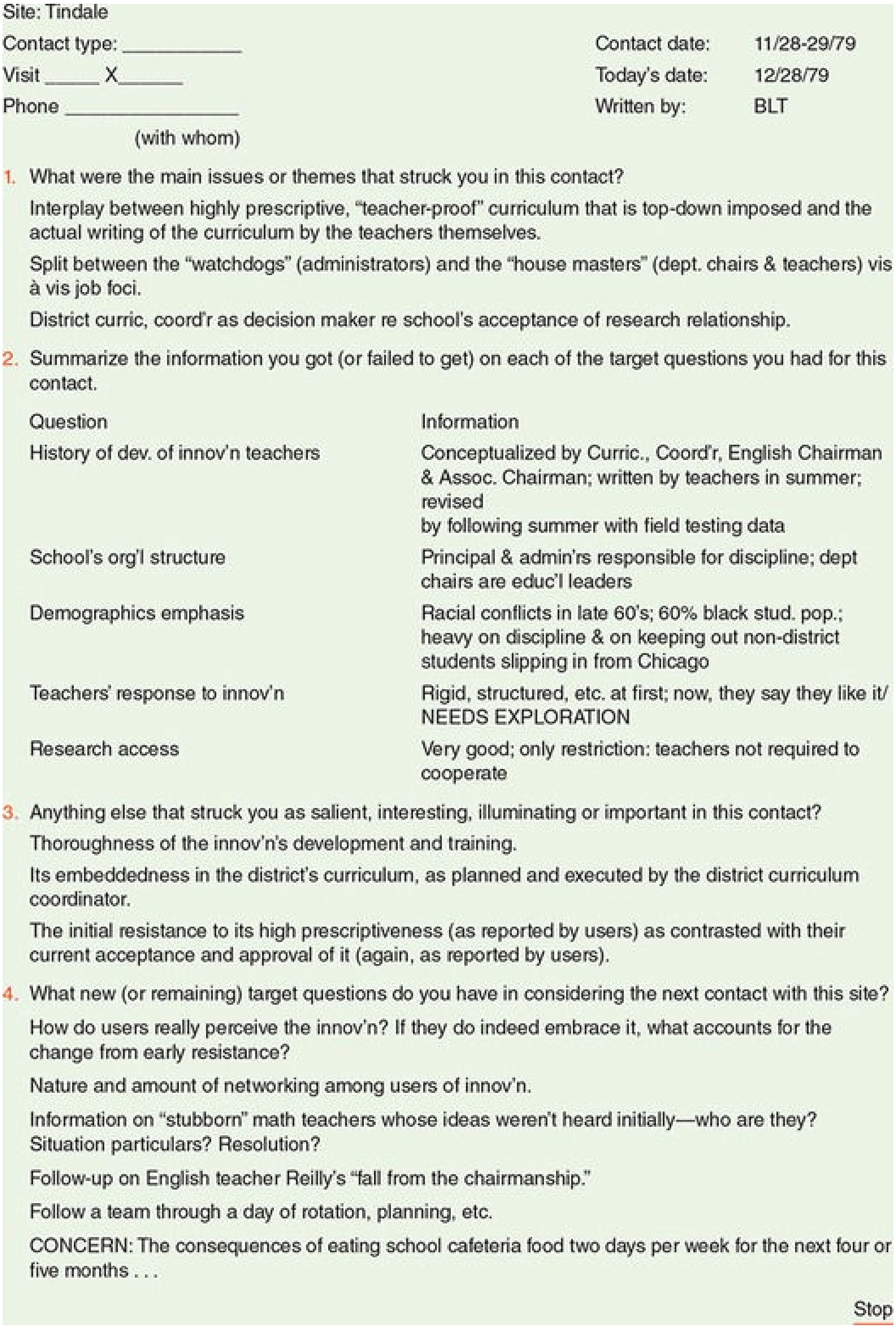
Documentation

The data for a qualitative study most often are notes jotted down in the field or during an interview—from which the original comments, observations, and feelings are reconstructed—or text transcribed from audio recordings. “The basic data are these observations and conversations, the actual words of people reproduced to the best of my ability from the field notes” (Diamond 1992:7). What to do with all this material? Many field research projects have slowed to a halt because a novice researcher becomes overwhelmed by the quantity of information that has been collected. A 1-hour interview can generate 20–25 pages of single-spaced text (Kvale 1996:169). Analysis is less daunting, however, if the researcher maintains a disciplined transcription schedule.

> Usually, I wrote these notes immediately after spending time in the setting or the next day. Through the exercise of writing up my field notes, with attention to “who” the speakers and actors were, I became aware of the nature of certain social relationships and their positional arrangements within the peer group. (Anderson 2003:235)

You can see the analysis already emerging from this simple process of taking notes.

Exhibit 11.3 Example of a Contact Summary Form



Site: Tindale

Contact type: ___________

Visit _____ X______

Phone ________________

(with whom)

Contact date: 11/28-29/79

Today's date: 12/28/79

Written by: BLT

1. What were the main issues or themes that struck you in this contact?

Interplay between highly prescriptive, "teacher-proof" curriculum that is top-down imposed and the actual writing of the curriculum by the teachers themselves.

Split between the "watchdogs" (administrators) and the "house masters" (dept. chairs & teachers) vis à vis job foci.

District curric, coord'r as decision maker re school's acceptance of research relationship.

2. Summarize the information you got (or failed to get) on each of the target questions you had for this contact.

Question	Information
History of dev. of innov'n teachers	Conceptualized by Curric., Coord'r, English Chairman & Assoc. Chairman; written by teachers in summer; revised by following summer with field testing data
School's org'l structure	Principal & admin'rs responsible for discipline; dept chairs are educ'l leaders
Demographics emphasis	Racial conflicts in late 60's; 60% black stud. pop.; heavy on discipline & on keeping out non-district students slipping in from Chicago
Teachers' response to innov'n	Rigid, structured, etc. at first; now, they say they like it/ NEEDS EXPLORATION
Research access	Very good; only restriction: teachers not required to cooperate

3. Anything else that struck you as salient, interesting, illuminating or important in this contact?

Thoroughness of the innov'n's development and training.

Its embeddedness in the district's curriculum, as planned and executed by the district curriculum coordinator.

The initial resistance to its high prescriptiveness (as reported by users) as contrasted with their current acceptance and approval of it (again, as reported by users).

4. What new (or remaining) target questions do you have in considering the next contact with this site?

How do users really perceive the innov'n? If they do indeed embrace it, what accounts for the change from early resistance?

Nature and amount of networking among users of innov'n.

Information on "stubborn" math teachers whose ideas weren't heard initially—who are they? Situation particulars? Resolution?

Follow-up on English teacher Reilly's "fall from the chairmanship."

Follow a team through a day of rotation, planning, etc.

CONCERN: The consequences of eating school cafeteria food two days per week for the next four or five months . . .

Stop

Source: Miles and Huberman (1994:10, Figure 4.1). Reprinted with permission from SAGE Publications, Inc.

The first formal analytic step is documentation. The various contacts, interviews, written documents, and whatever it is that preserves a record of what happened all need to be saved and listed. Documentation is critical to qualitative research for several reasons: It is essential for keeping track of what will be a rapidly growing volume of notes, audio and perhaps video recordings, and documents; it provides a way of developing and outlining the analytic process; and it encourages ongoing conceptualizing and strategizing about the text.

Matthew Miles and A. Michael Huberman (1994:53) provide a good example of a contact summary form that was used to keep track of observational sessions in a qualitative study of a new school curriculum (see Exhibit 11.3).

Organization, Categorization, and Condensation

Identifying and refining important concepts so that they can be organized and categorized is a key part of the iterative process of qualitative research. Sometimes, conceptual organization begins with a simple observation that is interpreted directly, “pulled apart,” and then put back together more meaningfully. Robert Stake (1995) provides an example:

> When Adam ran a push broom into the feet of the children nearby, I jumped to conclusions about his interactions with other children: aggressive, teasing, arresting. Of course, just a few minutes earlier I had seen him block the children climbing the steps in a similar moment of smiling bombast. So I was aggregating, and testing my unrealized hypotheses about what kind of kid he was, not postponing my interpreting. . . . My disposition was to keep my eyes on him. (p. 74)

The focus in this conceptualization “on the fly” is to provide a detailed description of what was observed and a sense of why that was important.

More often, analytic insights are tested against new observations, the initial statement of problems and concepts is refined, the researcher then collects more data, interacts with the data again, and the process continues. Anderson (2003) recounts how his conceptualization of social stratification at Jelly’s Bar developed over a long period:

> I could see the social pyramid, how certain guys would group themselves and say in effect, “I’m here and you’re there.” . . . I made sense of these crowds [initially] as the “respectables,” the “nonrespectables,” and the “near-respectables.” . . . Inside, such nonrespectables might sit on the crates, but if a respectable came along and wanted to sit there, the lower-status person would have to move. (pp. 225–226)

But this initial conceptualization changed with experience, as Anderson realized that the participants themselves used other terms to differentiate social status: *winehead, hoodlum,* and *regular* (Anderson 2003:230). What did they mean by these terms? The regulars basically valued “decency.” They associated decency with conventionality but also with “working for a living,” or having a “visible means of support” (Anderson 2003:231). In this way, Anderson progressively refined his concept as he gained experience in the setting.

Howard Becker (1958) provides another excellent illustration of this iterative process of conceptual organization in his study of medical students:

> When we first heard medical students apply the term “crock” to patients, we made an effort to learn precisely what they meant by it. We found, through interviewing students about cases both they and the observer had seen, that the term referred in a derogatory way to patients with many subjective symptoms but no discernible physical pathology. Subsequent observations indicated that this usage was a regular feature of student behavior and thus that we should attempt to incorporate this fact into our model of student-patient behavior. The derogatory character of the term suggested in particular that we investigate the reasons students disliked these patients. We found that this dislike was related to what we discovered to be the students’ perspective on medical school: the view that they were in school to get experience in recognizing and treating those common diseases most likely to be encountered in general practice. “Crocks,” presumably having no disease, could furnish no such experience. We were thus led to specify connections between the student-patient relationship and the student’s view of the purpose of this professional education. Questions concerning the genesis of this perspective led to discoveries about the organization of the student body and communication among students, phenomena which we had been assigning to another [segment of the larger theoretical model being developed]. Since “crocks” were also disliked because they gave the student no opportunity to assume medical responsibility, we were able to connect this aspect of the student-patient relationship with still another tentative model of the value system and hierarchical organization of the school, in which medical responsibility plays an important role. (p. 658)

This excerpt shows how the researcher first was alerted to a concept by observations in the field, then refined his understanding of this concept by investigating its meaning. By observing the concept’s frequency of use, he came to realize its importance. Then he incorporated the concept into an explanatory model of student–patient relationships.

A well-designed chart, or matrix, can facilitate the coding and categorization process. Exhibit 11.4 shows an example of a coding form designed by Miles and Huberman (1994:93–95) to represent the extent to which teachers and teachers’ aides (“users”) and administrators at a school gave evidence of various supporting conditions that indicate preparedness for a new reading program. The matrix condenses data into simple categories, reflects further analysis of the data to identify degree of support, and provides a multidimensional summary that will facilitate subsequent, more intensive analysis. Direct quotes still impart some of the flavor of the original text.

Matrix: A form on which particular features of multiple cases or instances can be recorded systematically so that a qualitative data analyst can examine them later.

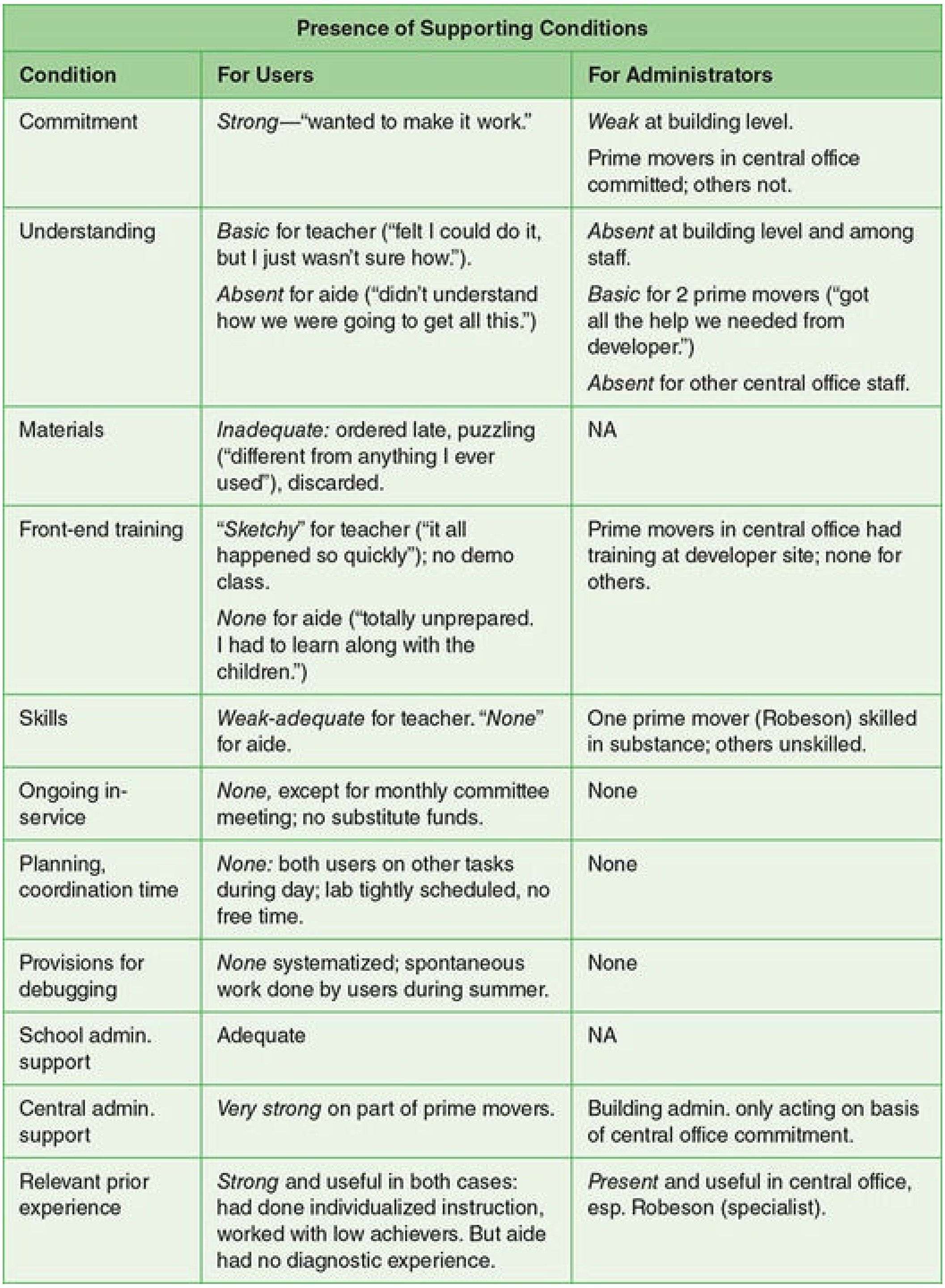
Examination and Display of Relationships

Examining relationships is the centerpiece of the analytic process because it allows the researcher to move from simple descriptions of the people and settings to explanations of why things happened as they did with those people in those settings. The process of examining relationships can be captured in a matrix that shows how different concepts are connected, or perhaps what causes are linked with what effects.

Exhibit 11.5 displays a matrix used to capture the relationship between the extent to which stakeholders in a new program had something important at stake in the program and the researcher's estimate of their favorability toward the program. Each cell of the matrix was to be filled in with a summary of an illustrative case study. In other matrix analyses, quotes might be included in the cells to represent the opinions of these different stakeholders, or the number of cases of each type might appear in the cells. The possibilities are almost endless. Keeping this approach in mind will generate many fruitful ideas for structuring a qualitative data analysis.

The simple relationships that are identified with a matrix like that shown in Exhibit 11.5 can be examined and then extended to create a more complex causal model. Such a model represents the multiple relationships between the constructs identified in a qualitative analysis as important for explaining some outcome. A great deal of analysis must precede the construction of such a model, with careful attention to identification of important variables and the evidence that suggests connections between them. Exhibit 11.6 provides an example of these connections from a study of the implementation of a school program.

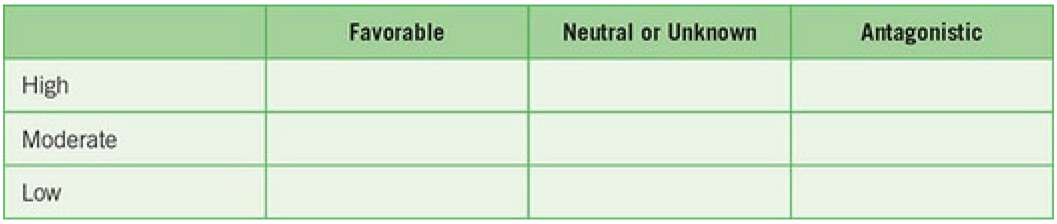
Exhibit 11.4 Example of a Checklist Matrix



Presence of Supporting Conditions		
Condition	**For Users**	**For Administrators**
Commitment	*Strong*—"wanted to make it work."	*Weak* at building level. Prime movers in central office committed; others not.
Understanding	*Basic* for teacher ("felt I could do it, but I just wasn't sure how."). *Absent* for aide ("didn't understand how we were going to get all this.")	*Absent* at building level and among staff. *Basic* for 2 prime movers ("got all the help we needed from developer.") *Absent* for other central office staff.
Materials	*Inadequate:* ordered late, puzzling ("different from anything I ever used"), discarded.	NA
Front-end training	"*Sketchy*" for teacher ("it all happened so quickly"); no demo class. *None* for aide ("totally unprepared. I had to learn along with the children.")	Prime movers in central office had training at developer site; none for others.
Skills	*Weak-adequate* for teacher. "*None*" for aide.	One prime mover (Robeson) skilled in substance; others unskilled.
Ongoing in-service	*None,* except for monthly committee meeting; no substitute funds.	None
Planning, coordination time	*None:* both users on other tasks during day; lab tightly scheduled, no free time.	None
Provisions for debugging	*None* systematized; spontaneous work done by users during summer.	None
School admin. support	Adequate	NA
Central admin. support	*Very strong* on part of prime movers.	Building admin. only acting on basis of central office commitment.
Relevant prior experience	*Strong* and useful in both cases: had done individualized instruction, worked with low achievers. But aide had no diagnostic experience.	*Present* and useful in central office, esp. Robeson (specialist).

Source: Miles and Huberman (1994:10, Table 5.2). Reprinted with permission from SAGE Publications, Inc.

Exhibit 11.5 Coding Form for Relationships: Stakeholders' Stakes



	Favorable	Neutral or Unknown	Antagonistic
High			
Moderate			
Low			

Source: Patton (2002).

Corroboration and Legitimization of Conclusions

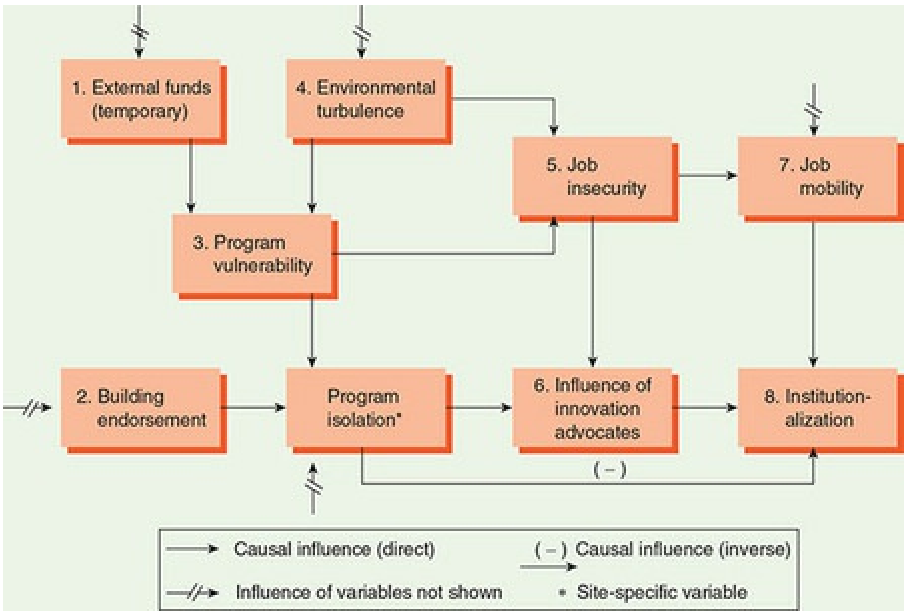
No set standards exist for evaluating the validity, or *authenticity,* of conclusions in a qualitative study, but the need to carefully consider the evidence and methods on which conclusions are based is just as great as with other types of research. Individual items of information can be assessed in terms of at least three criteria (Becker 1958):

1. *How credible was the informant?* Were statements made by someone with whom the researcher had a relationship of trust or by someone the researcher had just met? Did the informant have reason to lie? If the statements do not seem to be trustworthy as indicators of actual events, can they at least be used to help understand the informant's perspective?
2. *Were statements made in response to the researcher's questions, or were they spontaneous?* Spontaneous statements are more likely to indicate what would have been said had the researcher not been present.
3. *How does the presence or absence of the researcher or the researcher's informant influence the actions and statements of other group members?* Reactivity to being observed can never be ruled out as a possible explanation for some directly observed social phenomenon. However, if the researcher carefully compares what the informant says goes on when the researcher is not present, what the researcher observes directly, and what other group members say about their normal practices, the extent of reactivity can be assessed to some extent. (pp. 654–656)

A qualitative researcher's conclusions should also be assessed by his or her ability to provide a credible explanation for some aspect of social life. That explanation should capture group members' tacit knowledge of the social processes that were observed, not just their verbal statements about these processes. Tacit knowledge—"the largely unarticulated, contextual understanding that is often manifested in nods, silences, humor, and naughty nuances"—is reflected in participants' actions as well as their words and in what they fail to state but nonetheless feel deeply and even take for granted (Altheide and Johnson 1994:492–493). These features are evident in William Foote Whyte's (1955) analysis of Cornerville social patterns:

Tacit knowledge: In field research, a credible sense of understanding of social processes that reflects the researcher's awareness of participants' actions as well as their words, and of what they fail to state, feel deeply, and take for granted.

Exhibit 11.6 Example of a Causal Network Model


1. External funds (temporary)
4. Environmental turbulence
5. Job insecurity
7. Job mobility
3. Program vulnerability
2. Building endorsement
Program isolation*
6. Influence of innovation advocates
8. Institution-alization
(–)
Causal influence (direct)
(–) Causal influence (inverse)
Influence of variables not shown
* Site-specific variable

Source: Miles and Huberman (1994:10, Figure 6.5). Reprinted with permission from SAGE Publications, Inc.

> The corner-gang structure arises out of the habitual association of the members over a long period of time. The nuclei of most gangs can be traced back to early boyhood. . . . Home plays a very small role in the group activities of the corner boy. . . . The life of the corner boy proceeds along regular and narrowly circumscribed channels. . . . Out of [social interaction within the group] arises a system of mutual obligations which is fundamental to group cohesion. . . . The code of the corner boy requires him to help his friends when he can and to refrain from doing anything to harm them. When life in the group runs smoothly, the obligations binding members to one another are not explicitly recognized. (pp. 255–257)

Comparing conclusions from a qualitative research project to those obtained by other researchers while conducting similar projects can also increase confidence in their authenticity. Miller's (1999) study of NPOs (neighborhood police officers) found striking parallels in the ways they defined their masculinity to processes reported in research about males in nursing and other traditionally female jobs:

> In part, male NPOs construct an exaggerated masculinity so that they are not seen as feminine as they carry out the social-work functions of policing. Related to this is the almost defiant expression of heterosexuality, so that the men's sexual orientation can never truly be doubted even if their gender roles are contested. Male patrol officers' language—such as their use of terms like "pansy police" to

> connote neighborhood police officers—served to affirm their own heterosexuality. . . . In addition, the male officers, but not the women, deliberately wove their heterosexual status into conversations, explicitly mentioning their female domestic partner or spouse and their children. This finding is consistent with research conducted in the occupational field. The studies reveal that men in female-dominated occupations, such as teachers, librarians, and pediatricians, over-reference their heterosexual status to ensure that others will not think they are gay. (p. 222)

Reflection on the Researcher's Role

Confidence in the conclusions from a field research study is also strengthened by an honest and informative account about how the researcher interacted with subjects in the field, what problems he or she encountered, and how these problems were or were not resolved. Such a "natural history" of the development of the evidence enables others to evaluate the findings and reflects the constructivist philosophy that guides many qualitative researchers (see Chapter 1). Such an account is important primarily because of the evolving and variable nature of field research: To an important extent, the researcher "makes up" the method in the context of a particular investigation rather than applying standard procedures that are specified before the investigation begins.

Barrie Thorne (1993) provides a good example of this final element of the analysis:

> Many of my observations concern the workings of gender categories in social life. For example, I trace the evocation of gender in the organization of everyday interactions, and the shift from boys and girls as loose aggregations to "the boys" and "the girls" as self-aware, gender-based groups. In writing about these processes, I discovered that different angles of vision lurk within seemingly simple choices of language. How, for example, should one describe a group of children? A phrase like "six girls and three boys were chasing by the tires" already assumes the relevance of gender. An alternative description of the same event —"nine fourth-graders were chasing by the tires"—emphasizes age and downplays gender. Although I found no tidy solutions, I have tried to be thoughtful about such choices. . . . After several months of observing at Oceanside, I realized that my field notes were peppered with the words "child" and "children," but that the children themselves rarely used the term. "What do they call themselves?" I badgered in an entry in my field notes. The answer it turned out, is that children use the same practices as adults. They refer to one another by using given names ("Sally," "Jack") or language specific to a given context ("that guy on first base"). They rarely have occasion to use age-generic terms. But when pressed to locate themselves in an age-based way, my informants used "kids" rather than "children." (pp. 8–9)

Qualitative data analysts, more often than quantitative researchers, display real sensitivity to how a social situation or process is interpreted from a particular background and set of values and not simply based on the situation itself (Altheide and Johnson 1994). Researchers are only human, after all, and must rely on their own senses and process all information through their own minds. By reporting how and why they think they did what they did, they can help others determine whether, or how, the researchers' perspectives influenced their conclusions. “There should be clear ‘tracks’ indicating the attempt [to show the hand of the ethnographer] has been made” (Altheide and Johnson 1994:493).

Anderson’s (2003) memoir about the Jelly’s Bar research illustrates the type of “tracks” that an ethnographer makes as well as how the researcher can describe those tracks. Anderson acknowledges that his tracks began as a child:

> While growing up in the segregated black community of South Bend, from an early age, I was curious about the goings-on in the neighborhood, particularly the streets and more particularly the corner taverns where my uncles and my dad would go to hang out and drink. . . . Hence, my selection of a field setting was a matter of my background, intuition, reason, and a little bit of luck. (pp. 217–218)

After starting to observe at Jelly’s, Anderson’s (2003) tracks led to Herman:

> After spending a couple of weeks at Jelly’s, I met Herman. I felt that our meeting marked an important step. We would come to know each other well . . . something of an informal leader at Jelly’s. . . . We were becoming friends. . . . He seemed to genuinely like me, and he was one person I could feel comfortable with. (pp. 218–219)

So we learn that Anderson’s observations were to be shaped, in part, by Herman’s perspective, but we also find out that Anderson maintained some engagement with fellow students. This contact outside the bar helped shape his analysis: “By relating my experiences to my fellow students, I began to develop a coherent perspective, or a ‘story’ of the place that complemented the accounts I had detailed in my accumulating field notes” (Anderson 2003:220).

In this way, Anderson explains that the outcome of his analysis of qualitative data resulted, in part, from the way in which he “played his role” as a researcher and participant, not just from the setting itself.

Alternatives in Qualitative Data Analysis

The qualitative data analyst can choose from many interesting alternative approaches. Of course, the research question under investigation should shape the selection of an analytic approach, but the researcher's preferences and experiences also will inevitably have an important influence on the method chosen. The alternative approaches I present here—grounded theory, abductive analysis, case-oriented understanding, conversation analysis, narrative analysis, ethnomethodology, and qualitative comparative analysis—give you a good sense of the different possibilities (Patton 2002).

Grounded Theory

Theory development occurs continually in qualitative data analysis (Coffey and Atkinson 1996:23). Many qualitative researchers use a method of developing theory during their analysis that is termed grounded theory, which involves building up inductively a systematic theory that is *grounded* in, or based on, the observations. The grounded theorist first summarizes observations into conceptual categories and then tests the coherence of these categories directly in the research setting with more observations. Over time, as the researcher refines and links the conceptual categories, a theory evolves (Glaser and Strauss 1967; Huberman and Miles 1994:436). Exhibit 11.7 diagrams the grounded theory of a chronic illness "trajectory" developed by Anselm Strauss and Juliette Corbin (1990:221). Their notes suggested to them that conceptions of self, biography, and body are reintegrated after a process of grieving.

Grounded theory: A systematic theory developed inductively, based on observations that are summarized into conceptual categories, reevaluated in the research setting, and gradually refined and linked to other conceptual categories.

As observation, interviewing, and reflection continue, grounded theory researchers refine their definitions of problems and concepts and select indicators. They can then check the frequency and distribution of phenomena: How many people made a particular type of comment? How often did social interaction lead to arguments? Social system models may then be developed that specify the relationships between different phenomena. These models are modified as researchers gain experience in the setting. For the final analysis, the researchers check their models carefully against their notes and make a concerted attempt to discover negative evidence that might suggest that the model is incorrect.

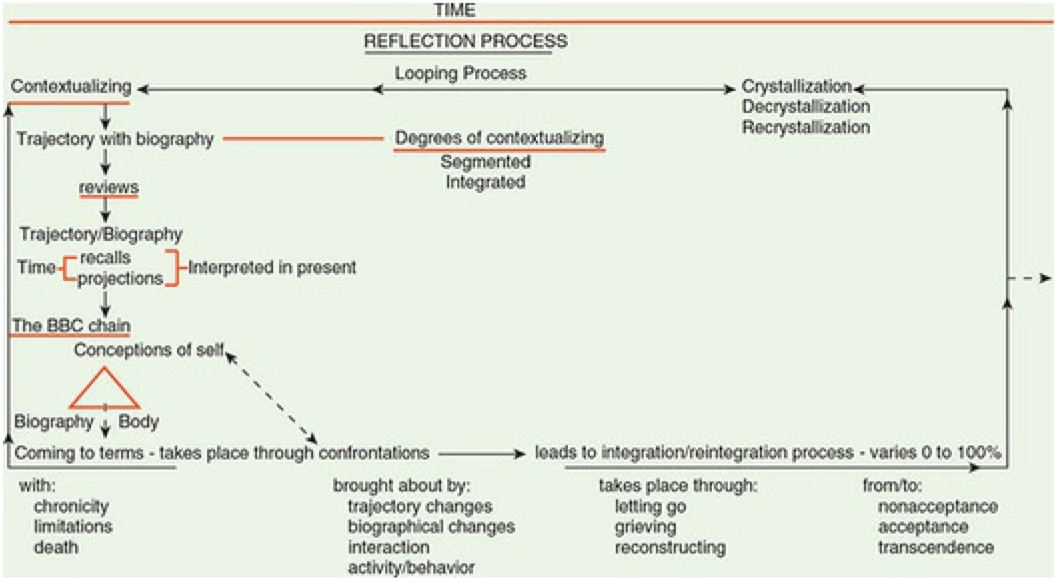
Heidi Levitt, Rebecca Todd Swanger, and Jenny Butler (2008:435) used a systematic grounded method of analysis to understand the perspective of male perpetrators of violence on female victims. Research participants were recruited from programs the courts used in Memphis to assess and treat perpetrators who admitted to having physically abused a female intimate partner. All program participants were of low socioeconomic status, but in other respects, Levitt and her colleagues (2008:436) sought to recruit a diverse sample.

The researchers (Levitt et al. 2008:437–438) began the analysis of their interview transcripts by dividing them into "meaning units"—"segments of texts that each contain one main idea"—and labeling these units with terms like those used by participants. The researchers then compared these labels and combined them into larger descriptive categories. This process continued until they had combined all the meaning units into seven different clusters. Exhibit 11.8 gives an example of two of their clusters and the four categories of meaning units combined within each (Levitt et al. 2008:439).

Here is how Levitt and her colleagues (2008) discuss the comments that were classified in Cluster 2, Category 3:

> Accordingly, when conflicts accumulated that could not be easily resolved, many of the men (5 of 12) thought that ending the relationship was the only way to stop violence from recurring. (p. 440)

Exhibit 11.7 A Grounded Theory Model



TIME
REFLECTION PROCESS
Looping Process
Contextualizing
Crystallization
Decrystallization
Recrystallization
Trajectory with biography
Degrees of contextualizing
Segmented
Integrated
reviews
Trajectory/Biography
Time
recalls
projections
Interpreted in present
The BBC chain
Conceptions of self
Biography
Body
Coming to terms - takes place through confrontations
leads to integration/reintegration process - varies 0 to 100%
with:
chronicity
limitations
death
brought about by:
trajectory changes
biographical changes
interaction
activity/behavior
takes place through:
letting go
grieving
reconstructing
from/to:
nonacceptance
acceptance
transcendence

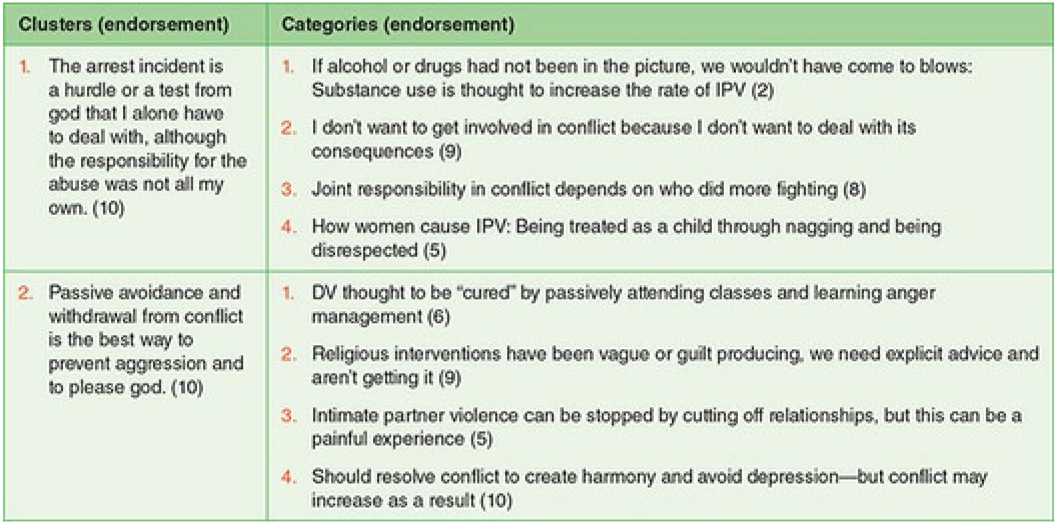
> "I don't deal with anybody so I don't have any conflicts. . . . It makes me feel bad because I be lonely sometime, but at the same time, it's the best thing going for me right now. I'm trying to rebuild me. I'm trying to put me on a foundation to where I can be a total leader. Like I teach my sons, 'Be leaders instead of followers.'" (p. 440)
>
> Although this interviewee's choice to isolate himself was a strategy to avoid relational dependency and conflict, it left him without interpersonal support and it could be difficult for him to model healthy relationships for his children. (p. 440)

With procedures such as these, the grounded theory approach develops general concepts from careful review of text or other qualitative materials and can then suggest plausible relationships between these concepts.

Abductive Analysis

But is it really possible to build up theory from observations that "speak for themselves"? Isn't the analyst always influenced by preexisting ideas, and aren't the qualitative researcher's observations going to be shaped in part by ideas about what it is important to observe? These are the questions that have led some researchers to question the logic of grounded theory and to propose an alternative approach to "theorizing qualitative research." After all, Iddo Tavory and Stefan Timmermans (2014:15) argue, induction "flows from theoretical frameworks that orient the analyst to a general framework of actions, meanings, institutional settings, and silences." What Tavory and Timmermans suggest is a more balanced approach—abductive analysis—that seeks unexpected and puzzling observations but recognizes that we need to interpret them in relation to theories that we already know. The analytic process is then more of a back-and-forth "ongoing conversation in which both scholars and research subjects take part, challenging our theorizations and suggesting new ones" (Tavory and Timmermans 2014:125, 130).

Exhibit 11.8 Clusters and Categories in a Grounded Theory Analysis



Clusters (endorsement)	Categories (endorsement)
1. The arrest incident is a hurdle or a test from god that I alone have to deal with, although the responsibility for the abuse was not all my own. (10)	1. If alcohol or drugs had not been in the picture, we wouldn't have come to blows: Substance use is thought to increase the rate of IPV (2) 2. I don't want to get involved in conflict because I don't want to deal with its consequences (9) 3. Joint responsibility in conflict depends on who did more fighting (8) 4. How women cause IPV: Being treated as a child through nagging and being disrespected (5)
2. Passive avoidance and withdrawal from conflict is the best way to prevent aggression and to please god. (10)	1. DV thought to be "cured" by passively attending classes and learning anger management (6) 2. Religious interventions have been vague or guilt producing, we need explicit advice and aren't getting it (9) 3. Intimate partner violence can be stopped by cutting off relationships, but this can be a painful experience (5) 4. Should resolve conflict to create harmony and avoid depression—but conflict may increase as a result (10)

Source: Levitt, H. M., Todd-Swanger, R., and Butler, J. B. "Male Perpetrators' Perspectives on Intimate Partner Violence, Religion, and Masculinity." Sex Roles: A Journal of Research, 58, 435–448. Copyright Ó 2007, Springer Science + Business Media, LLC. Reprinted with permission.

The difference between abductive analysis and grounded theory is in part a difference in how we view the generation of theory in relation to data: Is it possible to approach data about the world that we experience with a "blank slate" and then figure out what is going on in the setting we have studied (grounded theory) or not (abductive analysis)? But the difference also has to do with how we think of the social role of the qualitative data analyst.

In classical ethnography, the lone researcher goes “into the field” to gain experience without preconceptions, and then records and analyzes his or her observations in order to make sense of that field. In abductive analysis, the researcher prepares for gathering observations by reading a broad range of theories and, after collecting observations, engages actively in the scholarship and with the scholars and participants seeking to understand such settings (Tavory and Timmermans 2014:125, 131).

You don’t need to think of abductive analysis as a whole different approach to qualitative data analysis that you need to learn. Instead, recognize the problem of trying to approach qualitative data collection without preconceptions and consider orienting yourself to a setting for investigation by reading relevant theories of related social processes.

Abductive analysis: A qualitative data analysis approach in which researchers produce theoretical hunches for unexpected research findings and develop them with a systematic analysis of variation across a study in relation to a broad range of theories and interchange with other scholars.

Case-Oriented Understanding

Like many qualitative approaches, a case-oriented understanding attempts to understand a phenomenon from the standpoint of the participants. The case-oriented understanding method reflects an interpretive research philosophy that is not geared to identifying causes but provides a different way to explain social phenomena. For example, Constance Fischer and Frederick Wertz (2002) constructed such an explanation of the effect of being criminally victimized. They first recounted crime victims' stories and then identified common themes in these stories:

> Their explanation began with a description of what they termed the process of "living routinely" *before the crime: "he/she . . . feels that the defended against crime could never happen to him/her.*" "I said, 'nah, you've got to be kidding.'" (pp. 288–289, emphasis in original)
>
> In a second stage, "being disrupted," the victim copes with the discovered crime and fears worse outcomes: "You imagine the worst when it's happening . . . I just kept thinking my baby's upstairs." In a later stage, "reintegrating," the victim begins to assimilate the violation by taking some protective action: "But I clean out my purse now since then and I leave very little of that kind of stuff in there." (p. 289)
>
> Finally, when the victim is "going on," he or she reflects on the changes the crime produced: "I don't think it made me stronger. It made me smarter." (p. 290)

You can see how Fischer and Wertz (2002:288–290) constructed an explanation of the effect of crime on its victims through this analysis of the process of responding to the experience. This effort to "understand" what happened in these cases gives us a much better sense of why things happened as they did.

Case-oriented understanding: An understanding of social processes in a group, formal organization, community, or other collectivity that reflects accurately the standpoint of participants.

In the News Research in the News: How to Understand Solitary Confinement

For Further Thought?

During a decade of solitary confinement on death row in a Texas prison, Alfred D. Brown spent 22–24 hours in his 8′-×-12′ cell and sometimes an hour in a common room or outdoor courtyard, alone. His murder conviction was eventually thrown out due to evidence problems, but not before he was one of dozens of inmates who were interviewed as part of a qualitative study by the Human Rights Clinic at the University of Texas School of Law. The report's authors concluded that the practice of solitary confinement they studied was a form of torture; depriving human beings of any social contact is one of the worst possible forms of punishment.

1. The researchers were only allowed access to prisoners who had left death row. What do you think could be lost in interviews with those who were no longer confined in this way?
2. How would you approach analyzing interview data from prisoners with such experiences? What steps would you take to ensure that you captured the meanings they gave to their experiences? What evidence would you look for as you sought to construct a grounded theory about this topic?

News source: Fortin, Jack. 2017. "Report Compares Texas' Solitary Confinement Policies to Torture." *The New York Times,* April 26.

Conversation Analysis

Conversation analysis is a specific qualitative method for analyzing the sequential organization and details of conversation. Like ethnomethodology (described later), from which it developed, conversation analysis focuses on how reality is constructed, rather than on what it is. From this perspective, detailed analysis of conversational interaction is important because conversation is “sociological bedrock”: “a form of social organization through which the work of . . . institutions such as the economy, the polity, the family, socialization, etc.” is accomplished (Schegloff 1996:4).

> It is through conversation that we conduct the ordinary affairs of our lives. Our relationships with one another, and our sense of who we are to one another is generated, manifest, maintained, and managed in and through our conversations, whether face-to-face, on the telephone, or even by other electronic means. (Drew 2005:74)

Three premises guide conversation analysis (Gubrium and Holstein 2000:492):

1. Interaction is sequentially organized, and talk can be analyzed in terms of the process of social interaction rather than in terms of motives or social status.
2. Talk, as a process of social interaction, is contextually oriented—it is both shaped by interaction and creates the social context of that interaction.
3. These processes are involved in all social interaction, so no interactive details are irrelevant to understanding it.

Conversation analysis: A qualitative method for analyzing the sequential organization and details of ordinary conversation.

Consider these premises as you read the following excerpt from Elizabeth Stokoe’s (2006:479–480) analysis of the relevance of gender categories to “talk-in-interaction.” The dialogue in Exhibit 11.9 is between four first-year British psychology students who must write up a description of some photographs of people. Stokoe incorporates stills from the video recording of the interaction into her analysis of both the talk and embodied conduct in interaction. In typical conversation analysis style, the text is broken up into brief segments that capture shifts in meaning, changes in the speaker, pauses, nonspeech utterances and nonverbal actions, and emphases.

Can you see how the social interaction reinforces the link of “woman” and “secretary”?

Here, in part, is how Stokoe (2006) analyzes this conversation:

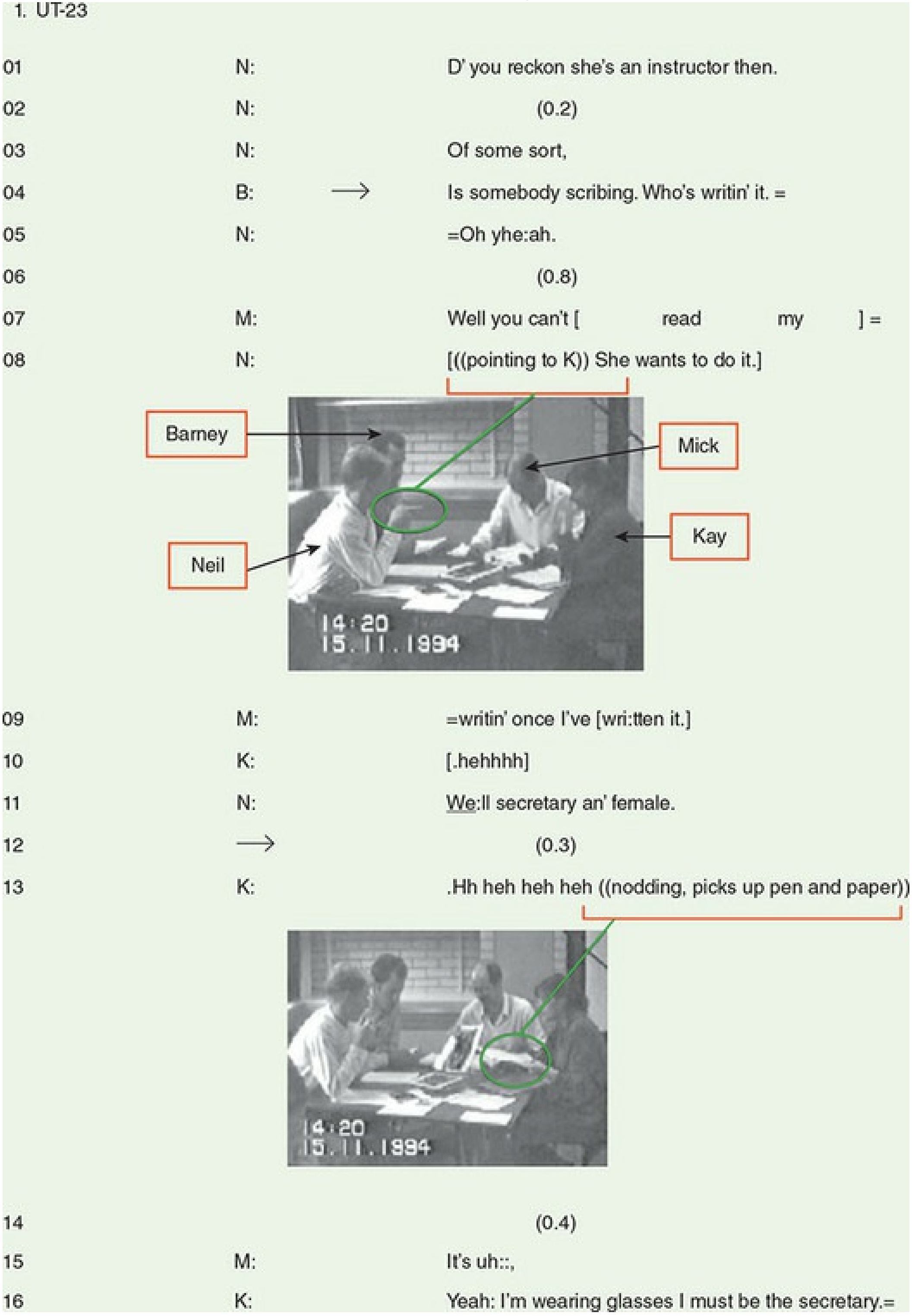
> In order to meet the task demands, one member of the group must write down their ideas. Barney's question at the start of the sequence, "is somebody scribing" is taken up after a reformulation: "who's writin' it." Note that, through a variety of strategies, members of the group manage their responses such that they do not have to take on the role of scribe. At line 05, Neil's "Oh yhe:ah." [the colon indicates prolongation of the phrase (Heritage n.d.:30)] treats Barney's turn as a proposal to be agreed with, rather than a request for action, and his subsequent nomination of Kay directs the role away from himself. . . . At line 08, Neil nominates Kay, his pointing gesture working in aggregate with the talk to accomplish the action ("She wants to do it."), whilst also attributing agency to Kay for taking up the role. A gloss [interpretation] might be "Secretaries in general are female, you're female, so you in particular are our secretary." (p. 481)

Bethan Benwell and Elizabeth Stokoe (2006:61–62) used a conversation between three friends to illustrate key concepts in conversation analysis. The text is prepared for analysis by numbering the lines, identifying the speakers, and inserting ↑ symbols to indicate inflection and decimal numbers to indicate elapsed time.

> 104 Marie: ↑ Has ↑ anyone-(0.2) has anyone got any really non:
>
> 105 sweaty stuff.
>
> 106 Dawn: Dave has, but you'll smell like a ma:n,
>
> 107 (0.9)
>
> 108 Kate: Eh [↑ huh heh]
>
> 109 Marie: [Right has] anyone got any ↑ fe:minine non sweaty stuff.

The gap at line 107, despite being less than a second long, is nevertheless quite a long time in conversation and indicates an interactional glitch or trouble. As Kate starts to laugh, Marie reformulates her request, from "↑ has ↑ anyone got any really non: sweaty stuff," to "right has anyone got any, ↑ fe:minine non sweaty stuff." The word *really* is replaced by *feminine* and is produced with an audible increase in pitch and emphasis. This replacement, together with the addition of *right,* displays her understanding of the problem with her previous question. For these speakers, smelling like a man (when one is a woman) is treated as a trouble source, a laughable thing and something that needs attending to and fixing. In this way, conversation analysis can uncover meanings in interactions about which the participants are not fully aware (Antaki 2008:438).

Exhibit 11.9 Conversation Analysis, Including Pictures


1. UT-23

01 N: D' you reckon she's an instructor then.
02 N: (0.2)
03 N: Of some sort,
04 B: → Is somebody scribing. Who's writin' it. =
05 N: =Oh yhe:ah.
06 (0.8)
07 M: Well you can't [read my] =
08 N: [((pointing to K)) She wants to do it.]
09 M: =writin' once I've [wri:tten it.]
10 K: [.hehhhh]
11 N: We:ll secretary an' female.
12 → (0.3)
13 K: .Hh heh heh heh ((nodding, picks up pen and paper))
14 (0.4)
15 M: It's uh::,
16 K: Yeah: I'm wearing glasses I must be the secretary.=

Source: Stokoe, Elizabeth. "On Ethnomethodology, Feminism, and the Analysis of Categorical Reference to Gender in Talk-In-Interaction." The Sociological Review 54:467–494.

Narrative Analysis

Narrative methods use interviews and sometimes documents or observations to “follow participants down their trails” (Riessman 2008:24). Unlike conversation analysis, which focuses attention on moment-by-moment interchange, narrative analysis seeks to put together the “big picture” about experiences or events as the participants understand them. Narrative analysis focuses on “the story itself” and seeks to preserve the integrity of personal biographies or a series of events that cannot adequately be understood in terms of their discrete elements (Riessman 2008:23–27). The analysis will often begin with a reconstruction of the chronology of a life or series of events, and then fill in the meanings attached to experiences throughout the chronology (Fawcett and Pockett 2015:59–62). Narrative thus “displays the goals and intentions of human actors; it makes individuals, cultures, societies, and historical epochs comprehensible as wholes” (Richardson 1995:200). The coding for a narrative analysis is typically of the narratives as a whole, rather than of the different elements within them. The coding strategy revolves around reading the stories and classifying them into general patterns.

> Narrative analysis: A form of qualitative analysis in which the analyst focuses on how respondents impose order on the flow of experience in their lives and thus make sense of events and actions in which they have participated.

For example, Morrill and his colleagues (2000:534) read through 254 conflict narratives written by the ninth graders they studied and found four different types of stories:

1. *Action tales,* in which the author represents himself or herself and others as acting within the parameters of taken-for-granted assumptions about what is expected for particular roles among peers.
2. *Expressive tales,* in which the author focuses on strong, negative emotional responses to someone who has wronged him or her.
3. *Moral tales,* in which the author recounts explicit norms that shaped his or her behavior in the story and influenced the behavior of others.
4. *Rational tales,* in which the author represents himself or herself as a rational decision maker navigating through the events of the story.

In addition to these dominant distinctions, Morrill et al. (2000:534–535) also distinguished the stories by four stylistic dimensions: (1) plot structure (e.g., whether the story unfolds sequentially), (2) dramatic tension (how the central conflict is represented), (3) dramatic resolution (how the central conflict is resolved), and (4) predominant outcomes (how the story ends). Coding reliability was checked through a discussion between the two primary coders, who found that their classifications agreed for a large percentage of the stories.

The excerpt that begins this chapter exemplifies what Morrill et al. (2000) termed an *action tale.* Such tales

> unfold in matter-of-fact tones kindled by dramatic tensions that begin with a disruption of the quotidian order of everyday routines. A shove, a bump, a look . . . triggers a response. . . . Authors of action tales typically organize their plots as linear streams of events as they move briskly through the story's scenes. . . . This story's dramatic tension finally resolves through physical fighting, but . . . only after an attempted conciliation. (p. 536)

You can contrast this action tale with the following narrative, which Morrill et al. (2000) classify as a *moral tale,* in which the students "explicitly tell about their moral reasoning, often referring to how normative commitments shape their decisionmaking" (p. 542):

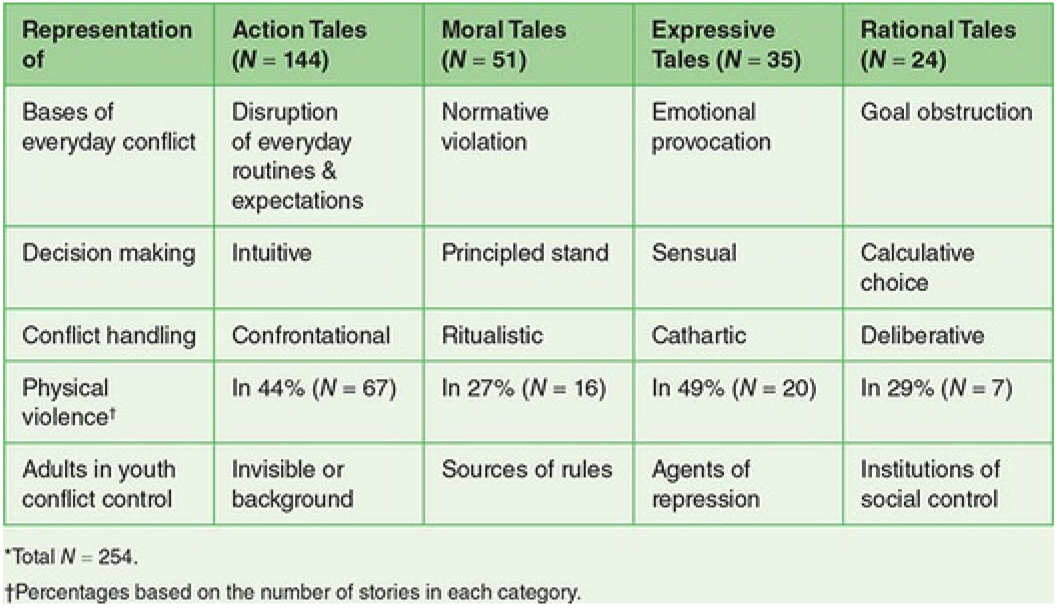
> I . . . got into a fight because I wasn't allowed into the basketball game. I was being harassed by the captains that wouldn't pick me and also many of the players. The same type of things had happened almost every day where they called me bad words so I decided to teach the ring leader a lesson. I've never been in a fight before but I realized that sometimes you have to make a stand against the people that constantly hurt you, especially emotionally. I hit him in the face a couple of times and I got [the] respect I finally deserved. (pp. 545–546)

Morrill et al. (2000:553) summarize their classification of the youth narratives in a simple table that highlights the frequency of each type of narrative and the characteristics associated with each of them (see Exhibit 11.10). How does such an analysis contribute to our understanding of youth violence? Morrill et al. (2000) first emphasize that their narratives "suggest that consciousness of conflict among youths—like that among adults—is not a singular entity, but comprises a rich and diverse range of perspectives" (p. 551).

Theorizing inductively, Morrill et al. (2000:553–554) then attempt to explain why action tales were much more common than were the more adult-oriented normative, rational, or emotionally expressive tales. One possibility is Carol Gilligan's (1988) theory of moral development, which suggests that younger students are likely to limit themselves to the simpler action tales that "concentrate on taken-for-granted assumptions of their peer and wider cultures, rather than on more self-consciously reflective interpretation and evaluation" (Morrill et al. 2000:554). More generally, Morrill et al. (2000) argue, "We can begin to think of the building blocks of cultures as different narrative styles in which various aspects of reality are accentuated, constituted, or challenged, just as others are deemphasized or silenced” (p. 556).

In this way, Morrill et al.’s (2000) narrative analysis allowed an understanding of youth conflict to emerge from the youths’ own stories while informing our understanding of broader social theories and processes.

Exhibit 11.10 Summary Comparison of Youth Narratives*



Representation of	Action Tales (*N* = 144)	Moral Tales (*N* = 51)	Expressive Tales (*N* = 35)	Rational Tales (*N* = 24)
Bases of everyday conflict	Disruption of everyday routines & expectations	Normative violation	Emotional provocation	Goal obstruction
Decision making	Intuitive	Principled stand	Sensual	Calculative choice
Conflict handling	Confrontational	Ritualistic	Cathartic	Deliberative
Physical violence†	In 44% (*N* = 67)	In 27% (*N* = 16)	In 49% (*N* = 20)	In 29% (*N* = 7)
Adults in youth conflict control	Invisible or background	Sources of rules	Agents of repression	Institutions of social control

*Total *N* = 254.

†Percentages based on the number of stories in each category.

Source: Morrill et al. 2000:553, Table 1, Copyright 2000. Reprinted with permission of Blackwell Publishing Ltd.

Narrative analysis can also use documents and observations and focus more attention on how stories are constructed, rather than on the resulting narrative (Hyvärinen 2008:452). Narrative analyst Catherine Kohler Riessman (2008:67–73) describes the effective combination of data from documents, interviews, and field observations to learn how members of Alcoholics Anonymous (AA) developed a group identity (Cain 1991). Propositions that Carol Cain (1991:228) identified repeatedly in the documents enter into stories as guidelines for describing the progression of drinking, the desire and inability to stop, the necessity of “hitting bottom” before the program can work, and the changes that take place in one’s life after joining AA.

Cain then found that this same narrative was expressed repeatedly in AA meetings. She interviewed only three AA members but found that one who had been sober and in AA for many years told “his story” using this basic narrative, while one who had been sober for only 2 years deviated from the narrative in some ways. One interviewee did not follow this standard narrative at all as he told his story; he had attended AA only sporadically for 20 years and left soon after the interview. Cain (1991) explains,

> I argue that as the AA member learns the AA story model, and learns to place the events and experiences of his own life into the model, he learns to tell and to understand his own life as an AA life, and himself as an AA alcoholic. The personal story is a cultural vehicle for identity acquisition. (p. 215)

Narrative inquiry is often a much more active process, involving purposeful sampling of participants in a setting and then interviews that the researcher uses to construct a narrative about their lives that may even be shared and refined with feedback from the participants (Edmonds and Kennedy 2016:163). In this way, Magnus Kilger (2017) interviewed elite youth sports players in Sweden to learn how they constructed "success stories." After recording 53 interviews during training sessions and matches, he read and interpreted each one, creating a classification scheme that included four types of narratives: *the humble story*—"we trained a bit" and "it just moved on"; *the hard work story*—"I'm fighting quite a lot and I'm [sic] have a strong desire"; *the natural talent story*—"I've always been, like, sort of the best player in the team"; and *the superstar story*—"I have [a] great number of very strong qualities . . . also an extreme speed . . . I'm very, very fast." In this way Kilger develops our understanding of how the players explain their success within their life stories.

Ethnomethodology

Ethnomethodology focuses on the way that participants construct the social world in which they live—how they "create reality"—rather than on describing the social world itself. Ethnomethodologists do not necessarily believe that we can find an objective reality; rather, the way that participants come to create and sustain a sense of reality is of interest. In the words of Jaber Gubrium and James Holstein (1997), in ethnomethodology, compared with the naturalistic orientation of ethnography (see Chapter 10),

> The focus shifts from the scenic features of everyday life onto the ways through which the world comes to be experienced as real, concrete, factual, and "out there." An interest in members' methods of constituting their world supersedes the naturalistic project of describing members' worlds as they know them. (p. 41)

Ethnomethodology: A qualitative research method focused on the way that participants in a social setting create and sustain a sense of reality.

Unlike the ethnographic analyst, who seeks to describe the social world as the participants see it, the ethnomethodological analyst seeks to maintain some distance from that world. The ethnomethodologist views a code of conduct like that described by Anderson (2003) at Jelly's not as a description of a real normative force that constrains social action, but as the way that people in the setting create a sense of order and social structure (Gubrium and Holstein 1997:44–45). The ethnomethodologist focuses on how reality is constructed, not on what it is.

Sociologist Harold Garfinkel (1967) developed ethnomethodology in the 1960s and first applied it to the study of gender. Focusing on a teenage male-to-female transsexual whom he termed "Agnes," he described her "social achievement of gender" as

> the tasks of securing and guaranteeing for herself the ascribed rights and obligations of an adult female by the acquisition and use of skills and capacities, the efficacious display of female appearances and performances, and the mobilizing of appropriate feelings and purposes. (p. 134)

The ethnomethodological focus on how the meaning of gender and other categories are socially constructed leads to a concern with verbal interaction and led to the more formal approach of conversation analysis (described earlier).

Qualitative Comparative Analysis

Daniel Cress and David Snow (2000) asked a series of very specific questions about social movement outcomes in their study of homeless social movement organizations (SMOs). They collected qualitative data from about 15 SMOs in eight cities. A content analysis of newspaper articles indicated that these cities represented a range of outcomes, and the SMOs within them were also relatively accessible to Cress and Snow because of prior contacts. In each of these cities, Cress and Snow used a snowball sampling strategy to identify the homeless SMOs and the various supporters, antagonists, and significant organizational bystanders with whom they interacted. Cress and Snow then gathered information from representatives of these organizations, including churches, other activist organizations, police departments, mayors' offices, service providers, federal agencies, and, of course, the SMOs themselves.

To answer their research questions, Cress and Snow (2000) needed to operationalize each of the various conditions that they believed might affect movement outcomes, using coding procedures that were much more systematic than those often employed in qualitative research. For example, Cress and Snow defined "sympathetic allies" operationally as

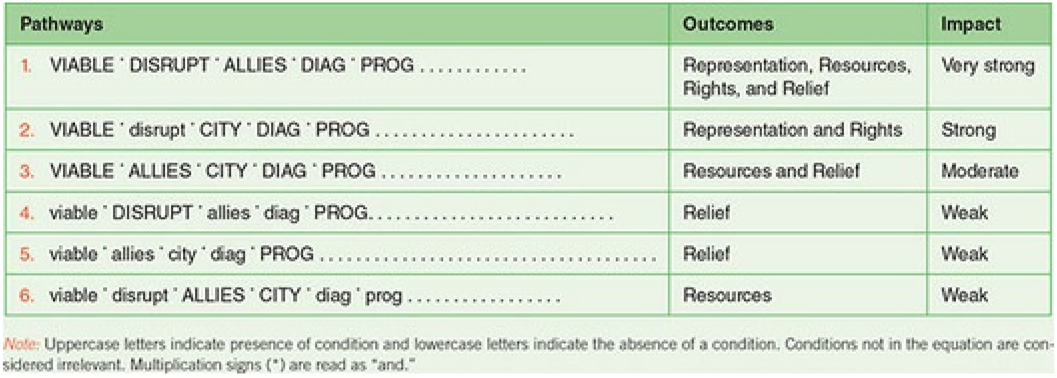
> the presence of one or more city council members who were supportive of local homeless mobilization. This was demonstrated by attending homeless SMO meetings and rallies and by taking initiatives to city agencies on behalf of the SMO. (Seven of the 14 SMOs had such allies.) (p. 1078)

Cress and Snow (2000) also chose a structured method of analysis, qualitative comparative analysis (QCA), to assess how the various conditions influenced SMO outcomes. This procedure identifies the combination of factors that had to be present across multiple cases to produce a particular outcome (Ragin 1987). Cress and Snow (2000) explain why QCA was appropriate for their analysis:

> QCA . . . is conjunctural in its logic, examining the various ways in which specified factors interact and combine with one another to yield particular outcomes. This increases the prospect of discerning diversity and identifying different pathways that lead to an outcome of interest and thus makes this mode of analysis especially applicable to situations with complex patterns of interaction among the specified conditions. (p. 1079)

Qualitative comparative analysis (QCA): A systematic type of qualitative analysis that identifies the combination of factors that had to be present across multiple cases to produce a particular outcome.

Exhibit 11.11 Multiple Pathways to Outcomes and Level of Impact



Pathways	Outcomes	Impact
1. VIABLE * DISRUPT * ALLIES * DIAG * PROG	Representation, Resources, Rights, and Relief	Very strong
2. VIABLE * disrupt * CITY * DIAG * PROG	Representation and Rights	Strong
3. VIABLE * ALLIES * CITY * DIAG * PROG	Resources and Relief	Moderate
4. viable * DISRUPT * allies * diag * PROG..........................	Relief	Weak
5. viable * allies * city * diag * PROG	Relief	Weak
6. viable * disrupt * ALLIES * CITY * diag * prog	Resources	Weak

Note: Uppercase letters indicate presence of condition and lowercase letters indicate the absence of a condition. Conditions not in the equation are considered irrelevant. Multiplication signs (*) are read as "and."

Source: Cress and Snow (2000:1097, Table 6). Copyright © 2000 The University of Chicago Press. Reprinted with permission.

Exhibit 11.11 summarizes the results of much of Cress and Snow's (2000) analysis. It shows that homeless SMOs that were coded as organizationally viable used disruptive tactics, had sympathetic political allies, and presented a coherent diagnosis and program in response to the problem they were protesting were very likely to achieve all four valued outcomes: (1) representation, (2) resources, (3) protection of basic rights, and (4) some form of tangible relief. Some other combinations of the conditions were associated with increased likelihood of achieving some valued outcomes, but most of these alternatives less frequently had positive effects.

The qualitative textual data on which the codes were based indicate how particular combinations of conditions exerted their influence. For example, one set of conditions that increased the likelihood of achieving increased protection of basic rights for homeless persons included avoiding disruptive tactics in cities that were more responsive to the SMOs. Cress and Snow (2000) use a quote from a local SMO leader to explain this process:

> We were going to set up a picket, but then we got calls from two people who were the co-chairs of the Board of Directors. They have like 200 restaurants. And they said, "Hey, we're not bad guys, can we sit down and talk?" We had been set on picketing. . . . Then we got to thinking, wouldn't it be better . . . if they co-drafted those things [rights guidelines] with us? So that's what we asked them to

> do. We had a work meeting, and we hammered out the guidelines. (p. 1089)

In Chapter 15, you will learn more about qualitative comparative analysis and see how this type of method can be used to understand political processes.

Combining Qualitative Methods

Qualitative researchers often combine one or more of these methods within one analysis. Elif Kale-Lostuvali (2007) enriched his research by using a combination of qualitative methodologies—including participant observation and intensive interviewing—to study the citizen–state encounters after the İzmit earthquake.

One important concept that emerged from both the observations and the interviews was the distinction between a *magǧdur* (sufferer) and a *depremzade* (son of the earthquake). This was a critical distinction, because a *magǧdur* was seen as deserving of government assistance, but a *depremzade* was considered to be taking advantage of the situation for personal gain. Kale-Lostuvali (2007) drew on both interviews and participant observation to develop an understanding of this complex concept:

> A prominent narrative that was told and retold in various versions all the time in the disaster area elaborated the contrast between *magǧdur* (sufferer; that is, the truly needy) and *depremzades* (sons of the earthquake) on the other. The *magǧdur* (sufferers) were the deserving recipients of the aid that was being distributed. However, they (1) were in great pain and could not pursue what they needed; or (2) were proud and could not speak of their need; or (3) were humble, always grateful for the little they got, and were certainly not after material gains; or (4) were characterized by a combination of the preceding. And because of these characteristics, they had not been receiving their rightful share of the aid and resources. In contrast, *depremzades* (sons of the earthquake) were people who took advantage of the situation. (p. 755)

The qualitative research by Spencer Moore and his colleagues (2004) on the social response to Hurricane Floyd demonstrates the interweaving of data from focus groups and from participant observation with relief workers:

> Reports of heroic acts by rescuers, innumerable accounts of "neighbors helping neighbors," and the comments of HWATF [task force] participants suggest that residents, stranded motorists, relief workers, and rescuers worked and came together in remarkable ways during the relief and response phases of the disaster.
>
> Like people get along better . . . they can talk to each other. People who hadn't talked before, they talk now, a lot closer. That goes, not only for the neighborhood, job-wise, organization-wise, and all that. . . . [Our] union sent some stuff for some of the families that were flooded out. (Focus Group #4) (pp.

210–211)

Analyses based on combining different qualitative methods in this way can yield a richer understanding of the social context under investigation.

Visual Sociology

For about 150 years, people have been creating a record of the social world with photography and more recently, with videos. This creates the possibility of "observing" the social world through photographs and films and of interpreting the resulting images as a "text." Some of the earliest U.S. sociologists included photographs in journal articles about social conditions, but the discipline turned away from visual representations by 1916 as part of a general effort to establish more scientific standards of evidence (Emmison, Smith, and Mayall 2012:23–24). Not until the 1970s did qualitative researchers Howard Becker (1974) and Erving Goffman (1979) again draw attention to the value of visual images in understanding social patterns. In more recent years, the availability of photos and videos has exploded because of the ease of taking them with smartphones and posting them on the Internet; already by 2012, almost half of Internet users had posted original photos or videos online (Rainie, Brenner, and Purcell 2012). Similarly, by 2016, 729 photos were being uploaded to Instagram every second and 125,406 videos were being viewed on YouTube (Liberatore 2016). As of December 2016, 5 billion photos had been uploaded and made publicly available at the photo-sharing Flickr site by its 51 million registered members, with about 1.68 million photos added per day (Michel 2012). As a result, increasing numbers of social scientists are collecting and analyzing visual representations of social life, and visual sociology has become a growth industry.

Visual sociology: Sociological research in which the social world is "observed" and interpreted through photographs, films, and other images.

Careers and Research

Laurel Person Mecca, MA, Assistant Director and Senior Research Specialist, Qualitative Data Analysis Program

Laurel Person Mecca was uncertain of the exact career she wanted to pursue during her graduate studies at the University of Pittsburgh. Then she happened upon the University Center for Social & Urban Research (UCSUR). It's hard to imagine a better place to launch a research career involving qualitative data analysis. Since 2005, the center has provided services and consultation to investigators in qualitative data analysis. Mecca used UCSUR to recruit participants for her own research and then made it clear to staff that she would love to work there after finishing her degree. Fourteen years later, she enjoys her work there more than ever.

One of the greatest rewards Mecca has found in her work is the excitement of discovering the unexpected when her preconceived notions about what research participants will tell her turn out to be incorrect. She also finds that her interactions with research participants provide a unique view into peoples' lives, thus providing insights in her own life and a richer understanding of the human condition. In addition to these personal benefits, Mecca has the satisfaction of seeing societal benefits from the projects she consults on, including the following: improving technologies designed to enhance independent living for elderly and disabled persons; exploring the barriers to participation in the Supplemental Nutrition Assistance Program (SNAP); and evaluating a program to improve parent–adolescent communication about sexual behaviors to reduce sexually transmitted infections and unintended teen pregnancies.

Mecca has some sound advice for students interested in careers that involve conducting research or using research results:

> Gain on-the-job experience while in college, even if it is an unpaid internship. Find researchers who are conducting studies that interest you, and inquire about working for them. Even if they are not posting an available position, they may bring you on board. Persistence pays off! You are much more likely to be selected for a position if you demonstrate a genuine interest in the work and if you continue to show your enthusiasm by following up.
>
> Definitely check out the National Science Foundation's (NSF) Research Experience for Undergraduates (REU) program. Though most of these internships are in the "hard" sciences, there are plenty of openings in social sciences disciplines. These internships include a stipend and, oftentimes, assistance with travel and housing. They are wonderful opportunities to work directly on a research project and may provide the additional benefit of a conference presentation and/or publication.

You have already seen in this chapter how Stokoe's conversation analysis of "gender talk" (2006) was enriched by her analysis of photographs. You also see later in this chapter how Robert Sampson and Stephen Raudenbush (1999) used systematic coding of videotaped observations to measure the extent of disorder in Chicago neighborhoods. Visual sociologists and other social researchers have been developing methods such as these to learn how others "see" the social world and to create images for further study. Continuous video recordings can help researchers unravel sequences of events and identify nonverbal expressions of feelings (Heath and Luff 2008:501). As in the analysis of written text, however, the visual sociologist must be sensitive to the way in which a photograph or film "constructs" the reality that it depicts.

The International Visual Sociology Association's (IVSA, http://visualsociology.org/about.html) statement of purpose identifies different ways in which visual images can be used in research.

Our Purpose

The purpose of IVSA is to promote the study, production, and use of imagery, visual data, and visually oriented materials in teaching, research, and applied activities. We also foster the use of still photography, film, video, and electronically transmitted images in sociology and other related fields. Together we work to encourage:

- documentary studies of everyday life in contemporary communities
- the interpretive analysis of art and popular visual representations of society
- studies about the social impact of advertising and the commercial use of images
- the analysis of archival images as sources of data on society and culture
- the study of the purpose and the meaning of image-making practices like recreational and family photography

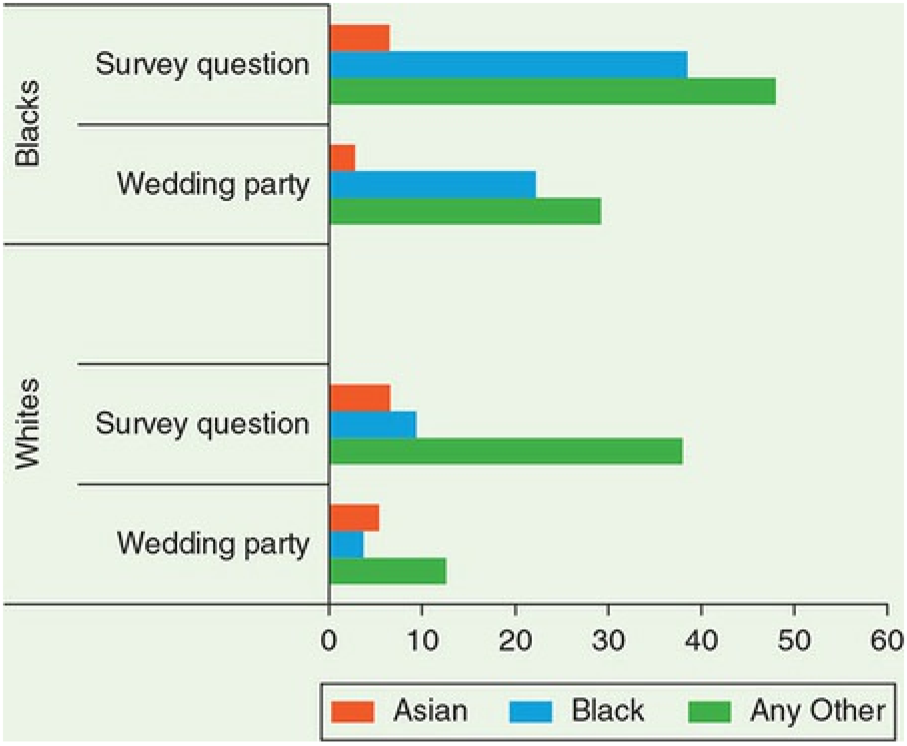
The research by Stokoe and by Sampson and Raudenbush both illustrate the first approach, the use of visual materials to document everyday life in contemporary communities. In both of these projects, the pictures were not the central method used, but they both extended the analysis of the other, quantitative data and gave additional insight into the social processes studied. In an innovative visually based approach to studying interracial friendship patterns, Brent Berry (2006) sampled wedding photos that had been posted on the web. Reasoning that bridesmaids and groomsmen represent who newlyweds consider to be their best friends, Berry compared the rate of different-race members of wedding parties to the prevalence of different-race friends reported in representative surveys. As you can see in Exhibit 11.12, answers to survey questions create the impression that interracial friendships are considerably more common than is indicated by the actual wedding party photos.

Alternatively, the researcher can analyze visual materials created by others (Emmison et al. 2012:20–21). Most of the purposes listed by the IVSA reflect this use of visual materials to learn about a society and culture.

An analysis by Eric Margolis (2004) of photographic representations of American Indian boarding schools provides a good example of how the analysis of visual materials created by others can help us understand cultural change. On the left side of Exhibit 11.13 is a picture taken in 1886 of Chiricahua Apaches who had just arrived at the Carlisle Indian School in Carlisle, Pennsylvania. The school was run by Captain Richard Pratt, who, like many Americans in that period, felt tribal societies were communistic, indolent, dirty, and ignorant, whereas Western civilization was industrious and individualistic. So Pratt set out to acculturate American Indians to the dominant culture. The second picture shows the result: the same group of Apaches looking like Europeans, not Native Americans—dressed in standard uniforms, with standard haircuts, and with more standard posture.

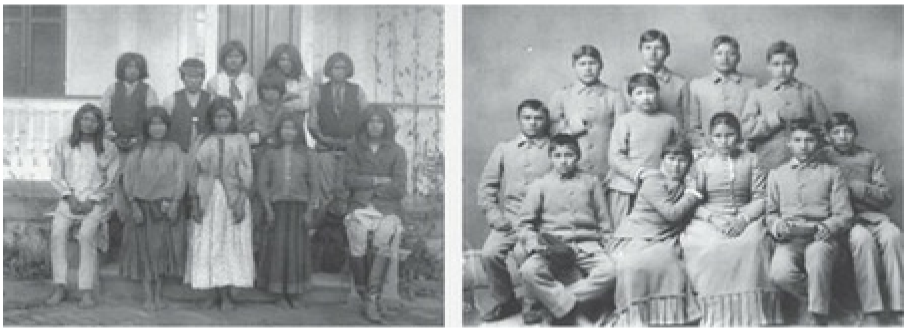
Many other pictures display the same type of transformation. Are these pictures each "worth a thousand words"? They capture the ideology of the school management, but we can be less certain that they document accurately the "before and after" status of the students. Pratt "consciously used photography to represent the boarding school mission as successful" (Margolis 2004:79). Although he clearly tried to ensure a high degree of conformity, there were accusations that the contrasting images were exaggerated to overemphasize the change (Margolis 2004:78). Reality was being constructed, not just depicted, in these photographs. Visual sociologists must always consider the purposes for which pictures were created and the extent to which people consciously posed for the occasion (Tinkler 2013:31). Even more important, visual representations must be analyzed in the context of the associated text and other indications of the social context of the photos. According to Luc Pauwels (2010:564), "It is important that visual researchers make every effort to situate the subject of their research and their specific take on it in its broader context, both visually and verbally."

Exhibit 11.12 Interracial Friendships in Wedding Party Photos and in Responses to Survey Questions



Source: Based on Berry, Brent. 2006. "Friends for Better or for Worse: Interracial Friendship in the United States as Seen Through Wedding Party Photos." Demography 43:491–510.

Exhibit 11.13 Pictures of Chiricahua Apache Children Before and After Starting Carlisle Indian School, Carlisle, Pennsylvania, 1886



Source: Margolis (2004:78).

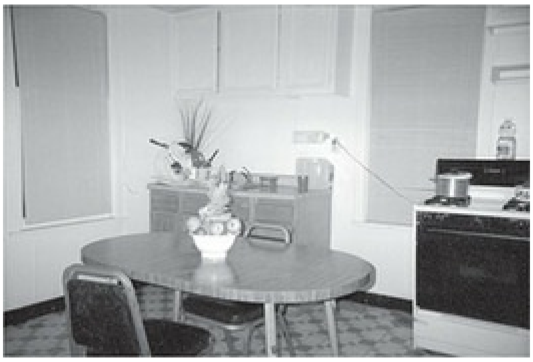
Darren Newbury (2005:1) cautioned the readers of his journal, *Visual Studies,* "images cannot be simply taken of the world, but have to be made within it." Reflecting this insight (and consistent with the last IVSA purpose), photo voice is a method of using photography to engage research participants in explaining how they have made sense of their social worlds. Rather than using images from other sources, the researcher directing a photo voice project distributes cameras to research participants and invites them to take pictures of their surroundings or everyday activities. The participants then meet with the researcher to present their pictures and discuss their meaning. In this way, researchers learn more about the participants' social worlds as they see it and react to it. The photo voice method also engages participants as part of the research team themselves, thus enriching the researcher's interpretations of the social world.

Photo voice: A method in which research participants take pictures of their everyday surroundings with cameras the researcher distributes, and then meet in a group with the researcher to discuss the pictures' meaning.

Lisa Frohmann (2005) recruited 42 Latina and South Asian women from battered women's support groups in Chicago to participate in research about the meaning of violence in their lives. Frohmann used photo voice methodology, so she gave each participant a camera. After they received some preliminary instruction, Frohmann invited participants to take about five to seven pictures weekly for 4–5 weeks. The photographs were to capture persons, places, and objects that represent the continuums of comfort–discomfort, happiness–sadness, safety–danger, security–vulnerability, serenity–anxiety, protection–exposure, strength–weakness, and love–hate (see Exhibit 11.14). Twenty-nine women then returned to discuss the results.

With this very simple picture, one participant, Jenny, described how family violence affected her feelings:

Exhibit 11.14 Picture in Photo Voice Project



Source: Frohmann (2005:1407).

> This is the dining room table and I took this picture because the table is empty and I feel that although I am with my children, I feel that it is empty because there is no family harmony, which I think is the most important thing. (Frohmann 2005:1407)

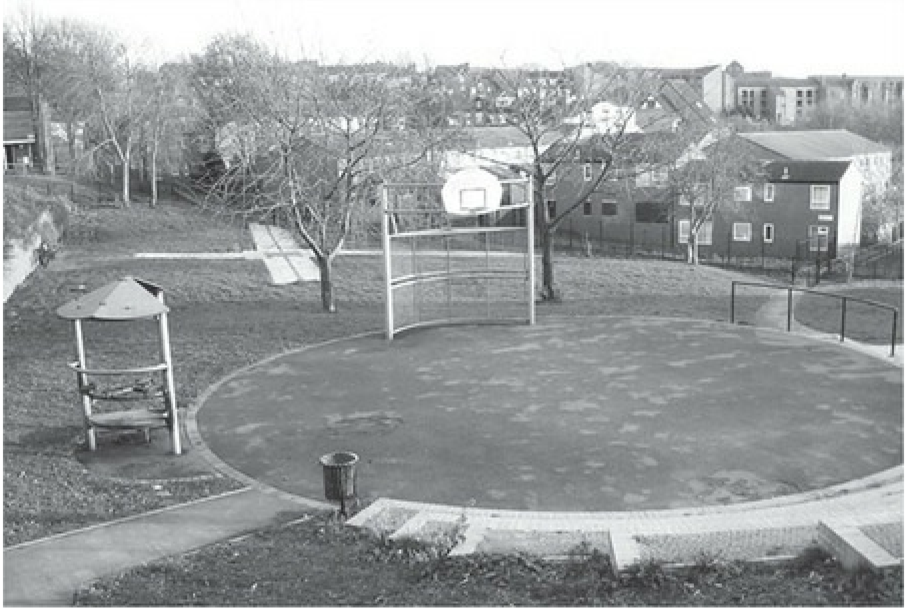
The image and narrative indirectly represent Jenny's concept of family: a husband and wife who love each other and their children. Food and eating together are important family activities. Part of caring for her family is preparing food. The photo shows that her concept of family is fractured (Frohmann 2005:1407).

Analysis of visual materials can also be used to enrich data collected with other methods. U.K. researchers Nick Emmel and Andrew Clark (2011) discuss how photographs collected in "walkarounds" enriched their understanding of the social setting they studied:

> The research is situated in one geographical location or fieldsite. Periodically we walked through this field along a set pathway taking photographs. . . . The research is conducted in a geographical place covering around 1.5 mile^2 (circa 2.5 km^2) with a mixed population. Relatively affluent students live in close proximity to one of the most deprived populations in England. . . . Within this socially heterogeneous geographical context our research explores, among other aims, the ways different social groups create, maintain, dissemble and experience, social networks over time and across space.
>
> We each use the photographs we take on the walk as an adjunct to the other methods we are using in the research. . . . . They contribute to and facilitate an interpretation of place, which in turn provides a more complete account of the

> place and space in which we are doing research.
>
> . . . how this analytical process happens. The panorama [see Exhibit 11.15] could be analysed at face value as an empty play area; perhaps supporting ideas about the out-migration of families (a common theme discussed by some resident groups). . . . Subsequent questioning about play spaces in the area however, reveals a range of alternative explanations for under-use. For example, conversational interviews with young people reveal a more nuanced geography of play and socialisation in the area; informal discussion with a local official suggest [*sic*] infrastructural problems with this particular space, while analysis of the recent history of this play space hints at a more political explanation for its existence and apparent under-use. This means that I do not analyse the images alone (that is, as a discrete data set); but rather alongside other methods. . . . Finally, . . . I use the walkaround method as a way of formulating new questions to ask of participants in the other methods. In some respects, it is the making of the photograph (deciding whether, and what, to photograph and why), rather than the image itself, that is more analytically revealing (pp. 2–3, 8–10).

Exhibit 11.15 A Playground in the Fieldsite of Emmel and Clark



Source: Emmel and Clark (2011:11).

The ease with which video can be captured with smartphones is also increasing the use of video in ethnographic projects, and is drawing attention to the method of video ethnography. This method combines the ethnographic process of engaging in ongoing social processes in a defined social setting with the visual sociologist’s attention to the nonverbal aspects of the social world. Although professional video ethnographic projects require use of a dedicated video camera, such as a handycam with “shotgun” microphone mounted on it that records sound directly in front of the camera (Shrum and Scott 2017:52–53), you can develop some skill in the method with just your smartphone.

With video gear in hand, a video ethnographer needs to consider the boundaries around the social processes of interest and what to film to capture those processes. It’s best seen as a component of an ethnographic study, rather than as a method in itself, for the video ethnographer has to “be there” when key interactions happen (Shrum and Scott 2017:28). You’ll inevitably miss a lot if you haven’t developed relations with actors in the setting who can help you understand the importance of events and be ready for critical moments. Of course, recording with a camera injects a new element into most social processes, but the widespread use of cameras in smartphones means that the presence of an ethnographer may be more disruptive than the fact of using a camera. A useful final product—a video that “tells a story”—will only emerge after extensive editing. The video editing process is comparable to the process of coding and selecting pictures and notes in a visual sociology or ethnography project, but will require specialized practice (Shrum and Scott 2017:93). You can even submit your video to withoutabox.com for consideration by film festival organizers!

Video ethnography: The use of audiovisual methods and editing techniques to record, analyze, and present one or more viewable social processes, actions, or events in interpretable segments.

Systematic Observation

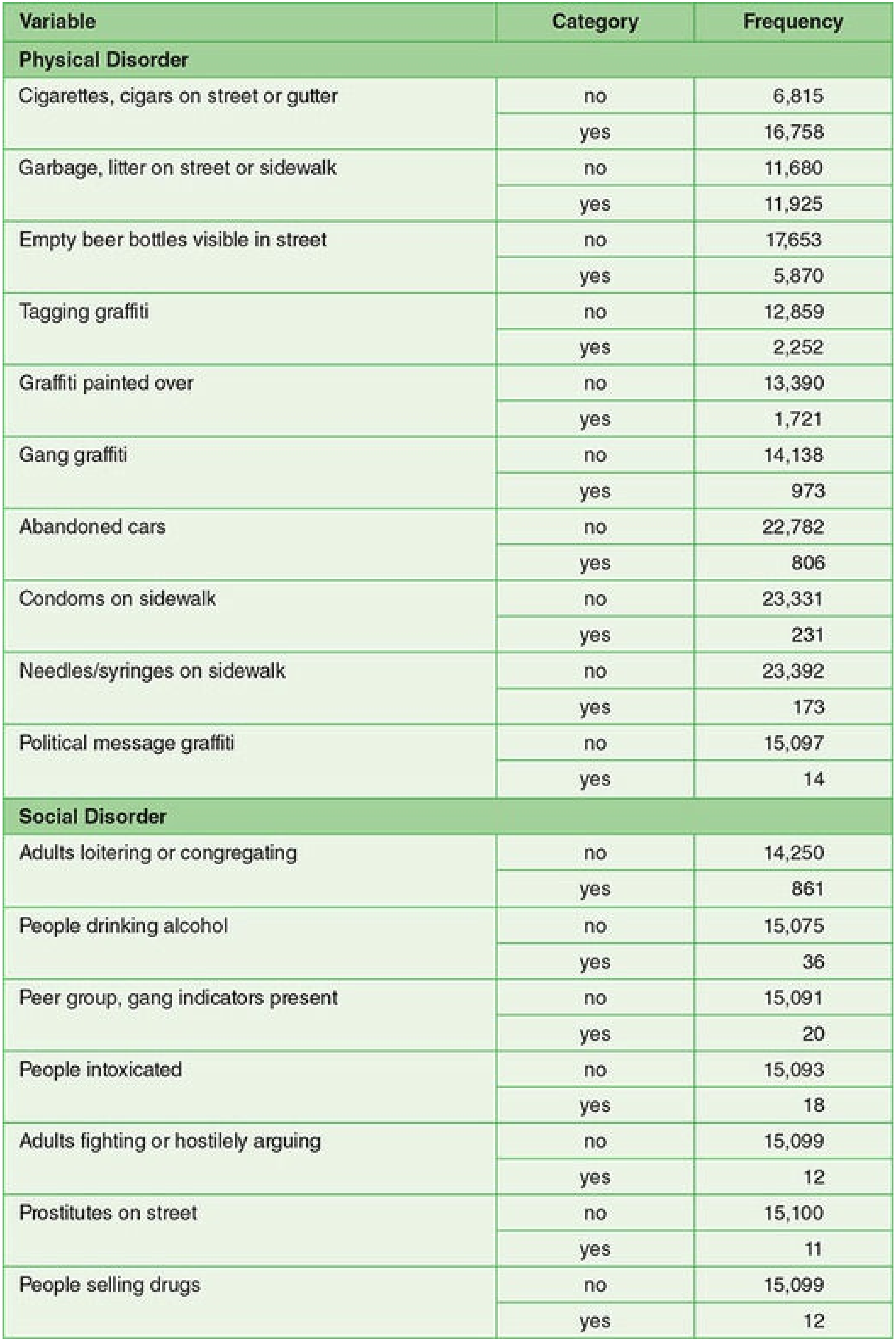
Observations can be made in a more systematic, quantitative design that allows systematic comparisons and more confident generalizations. A researcher using systematic observation develops a standard form on which to record variation within the observed setting in terms of variables of interest. Such variables might include the frequency of some behavior(s), the particular people observed, the weather or other environmental conditions, and the number and state of repair of physical structures. In some systematic observation studies, records will be obtained from a random sample of places or times.

Sampson and Raudenbush's (1999) study of disorder and crime in urban neighborhoods provides an excellent example of systematic observation methods. Although you learned about some features of this pathbreaking research in Chapter 6, in this section I elaborate on their use of systematic social observation to learn about these neighborhoods. A systematic observational strategy increases the reliability of observational data by using explicit rules that standardize coding practices across observers (Reiss 1971b). It is a method particularly well suited to overcome one of the limitations of survey research on crime and disorder: Residents who are fearful of crime perceive more neighborhood disorder than do residents who are less fearful, even though both are observing the same neighborhood (Sampson and Raudenbush 1999:606).

Systematic observation: A strategy that increases the reliability of observational data by using explicit rules that standardize coding practices across observers.

This ambitious multiple-methods investigation combined observational research, survey research, and archival research. The observational component involved a stratified probability (random) sample of 196 Chicago census tracts. A specially equipped sport-utility vehicle was driven down each street in these tracts at the rate of 5 miles per hour. Two video recorders taped the blocks on both sides of the street, while two observers peered out of the vehicle's windows and recorded their observations in the logs. The result was an observational record of 23,816 face blocks (the block on one side of the street is a face block). The observers recorded in their logs codes that indicated land use, traffic, physical conditions, and evidence of physical disorder (see Exhibit 11.16). The videotapes were sampled and then coded for 126 variables, including housing characteristics, businesses, and social interactions. Physical disorder was measured by counting such features as cigarettes or cigars in the street, garbage, empty beer bottles, graffiti, condoms, and syringes. Indicators of social disorder included adults loitering, drinking alcohol in public, fighting, and selling drugs. To check for reliability, a different set of coders recoded the videos for 10% of the blocks. The repeat codes achieved 98% agreement with the original codes.

Exhibit 11.16 Neighborhood Disorder Indicators Used in Systematic Observation Log



Variable	Category	Frequency
Physical Disorder		
Cigarettes, cigars on street or gutter	no	6,815
	yes	16,758
Garbage, litter on street or sidewalk	no	11,680
	yes	11,925
Empty beer bottles visible in street	no	17,653
	yes	5,870
Tagging graffiti	no	12,859
	yes	2,252
Graffiti painted over	no	13,390
	yes	1,721
Gang graffiti	no	14,138
	yes	973
Abandoned cars	no	22,782
	yes	806
Condoms on sidewalk	no	23,331
	yes	231
Needles/syringes on sidewalk	no	23,392
	yes	173
Political message graffiti	no	15,097
	yes	14
Social Disorder		
Adults loitering or congregating	no	14,250
	yes	861
People drinking alcohol	no	15,075
	yes	36
Peer group, gang indicators present	no	15,091
	yes	20
People intoxicated	no	15,093
	yes	18
Adults fighting or hostilely arguing	no	15,099
	yes	12
Prostitutes on street	no	15,100
	yes	11
People selling drugs	no	15,099
	yes	12

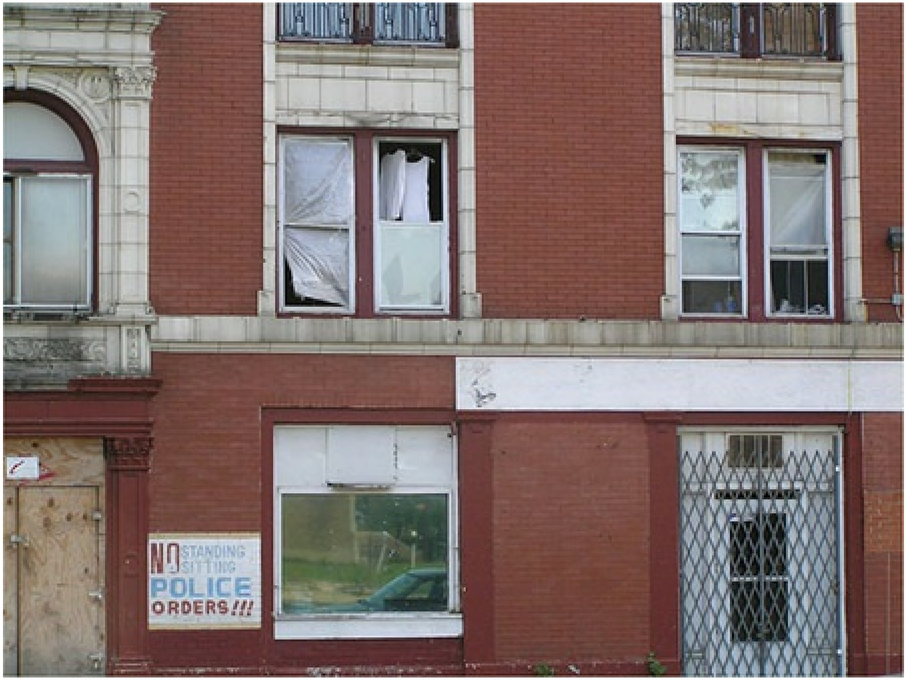
Source: Raudenbush and Sampson (1999:15).

Sampson and Raudenbush also measured crime levels with data from police records, census tract socioeconomic characteristics with census data, and resident attitudes and behavior with a survey. As you learned in Chapter 6, the combination of data from these sources allowed a test of the relative impact on the crime rate of informal social control efforts by residents and of the appearance of social and physical disorder.

Peter St. Jean (2007) extended the research of Sampson and Raudenbush with a mixed-methods study of high crime areas that used resident surveys, participant observation, in-depth interviews with residents and offenders, and systematic social observation. St. Jean recorded neighborhood physical and social appearances with video cameras mounted in a van that was driven along neighborhood streets. Pictures were then coded for the presence of neighborhood disorder (see Exhibit 11.17 and the study site for this book).

This study illustrates both the value of multiple methods and the technique of recording observations in a form from which quantitative data can be obtained. The systematic observations give us much greater confidence in the measurement of relative neighborhood disorder than we would have from unstructured descriptive reports or from responses of residents to survey questions. Interviews with residents and participant observation helped identify the reasons that offenders chose particular locations when deciding where to commit crimes.

Exhibit 11.17 One Building in St. Jean's (2007) Study



Source: © Peter K. B. St. Jean. Reprinted with permission.

Participatory Action Research

Whyte (1991) urged social researchers to engage with research participants throughout the research process. He formalized this recommendation into an approach he termed participatory action research (PAR). As the name implies, this approach encourages social researchers to get "out of the academic rut" and bring values into the research process (p. 285). Since Whyte's early call for this type of research, with a focus on research with organizational employees, PAR has become increasingly popular in disciplines ranging from public health to social work, as well as sociology (McIntyre 2008; Minkler 2000). Participatory action research is not itself a qualitative method, but PAR projects tend to use qualitative methods, which are more accessible to members of the lay public and which typically involve some of the same activities as in PAR: engaging with individuals in their natural settings and listening to them in their own words.

Participatory action research (PAR): A type of research in which the researcher involves members of the population to be studied as active participants throughout the research process, from the selection of a research focus to the reporting of research results and efforts to make changes based on the research; also termed *community-based participatory research.*

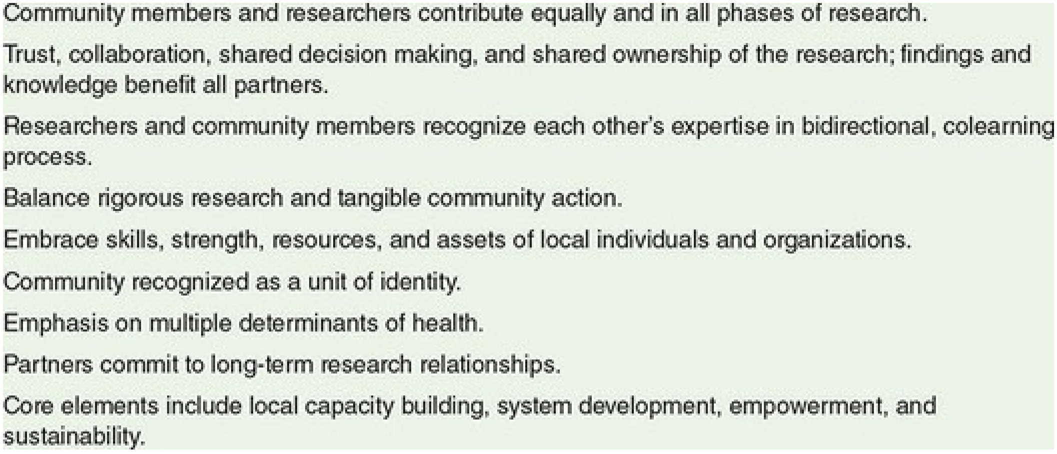
In PAR, also termed *community-based participatory research* (CBPR), the researcher involves as active participants some members of the setting studied. "The goal of CBPR is to create an effective translational process that will increase bidirectional connections between academics and the communities that they study" (Hacker 2013:2). Both the members and the researcher are assumed to want to develop valid conclusions, to bring unique insights, and to desire change, but Whyte (1991) believed that these objectives were more likely to be obtained if the researcher collaborates actively with the persons being studied. For example, many academic studies have found that employee participation is associated with job satisfaction but not with employee productivity. After some discussion about this finding with employees and managers, Whyte realized that researchers had been using a general concept of employee participation that did not distinguish those aspects of participation that were most likely to influence productivity (pp. 278–279). For example, occasional employee participation in company meetings had not been distinguished from ongoing employee participation in and control of production decisions. When these and other concepts were defined more precisely, it became clear that employee participation in production decisions had substantially increased overall productivity, whereas simple meeting attendance had not. This discovery would not have occurred without the active involvement of company employees in planning the research.

Those who engage in PAR projects are making a commitment "to listen to, learn from, solicit and respect the contributions of, and share power, information, and credit for accomplishments with the groups that they are trying [to] learn about and help" (Horowitz, Robinson, and Seifer 2009:2634). The emphasis on developing a partnership with the community is reflected in the characteristics of the method presented in a leading health journal (see Exhibit 11.18). Each characteristic (with the exception of "emphasis on multiple determinants of health") identifies a feature of the researcher's relationship with community members.

PAR can bring researchers into closer contact with participants in the research setting through groups that discuss and plan research steps and then take steps to implement research findings. For this reason, PAR is "particularly useful for emergent problems for which community partners are in search of solutions but evidence is lacking" (Hacker 2013:8). Stephen Kemmis and Robin McTaggart (2005:563–568) summarize the key steps in the process of conducting a PAR project as creating "a spiral of self-reflecting cycles":

- Planning a change
- Acting and observing the process and consequences of the change
- Reflecting on these processes and consequences
- Replanning
- Acting and observing again

Exhibit 11.18 Characteristics of Community-Based Participatory Research (CBPR)



Community members and researchers contribute equally and in all phases of research.

Trust, collaboration, shared decision making, and shared ownership of the research; findings and knowledge benefit all partners.

Researchers and community members recognize each other's expertise in bidirectional, colearning process.

Balance rigorous research and tangible community action.

Embrace skills, strength, resources, and assets of local individuals and organizations.

Community recognized as a unit of identity.

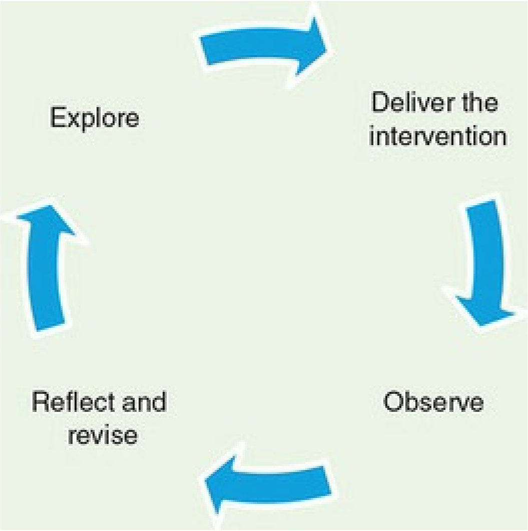
Emphasis on multiple determinants of health.

Partners commit to long-term research relationships.

Core elements include local capacity building, system development, empowerment, and sustainability.

Source: Horowitz, Carol R., Mimsie Robinson, and Sarena Seifer. 2009. "Community-Based Participatory Research From the Margin to the Mainstream: Are Researchers Prepared?" Circulation May 19, 2009. Table 1, Characteristics of CBPR, p. 2634. Reprinted with permission of Wolters Kluwer Health, LWW.

Exhibit 11.19 The Cyclical Action Research Approach



Source: Edmonds, W. Alex and Thomas D. Kennedy. 2016. An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods, 2nd ed, Thousand Oaks, CA: SAGE.

In contrast to the formal reporting of results at the end of a research project, these cycles make research reporting an ongoing part of the research process. “Community partners can deepen the interpretation process once results are available, as they are intimately familiar with the context and meaning” (Hacker 2013:8). Community partners may also work with the academic researchers to make changes in the community reflecting the research findings. Publication of results is only part of this cyclical research approach (see Exhibit 11.19).

Karen Hacker, while at Harvard University and the Institute for Community Health in Cambridge, Massachusetts, collaborated with community partners in response to a public health emergency in the adjacent town of Somerville (Hacker 2013:8–10; Hacker et al. 2008). After a series of youth suicides and overdoses from 2000 to 2005, a PAR community coalition was formed with members from mental health service providers, school leaders, police, and community parents. After reviewing multiple statistics, the coalition concluded that the deaths represented a considerable increase over previous years. However, when mental health professionals attempted to interview family members of adolescents who had died by suicide to investigate the background to the suicides, they were rebuffed; in contrast, family members were willing to talk at length with PAR members from their community. The PAR team was then able to map the relationships between the adolescents. The process of using the results of this research to respond to the suicides included a candlelight vigil, a speak-out against substance abuse, the provision of crisis counseling, and programs to support families and educate the community. Subsequently, the suicide rate dropped back to its pre-2000 level.

Computer-Assisted Qualitative Data Analysis

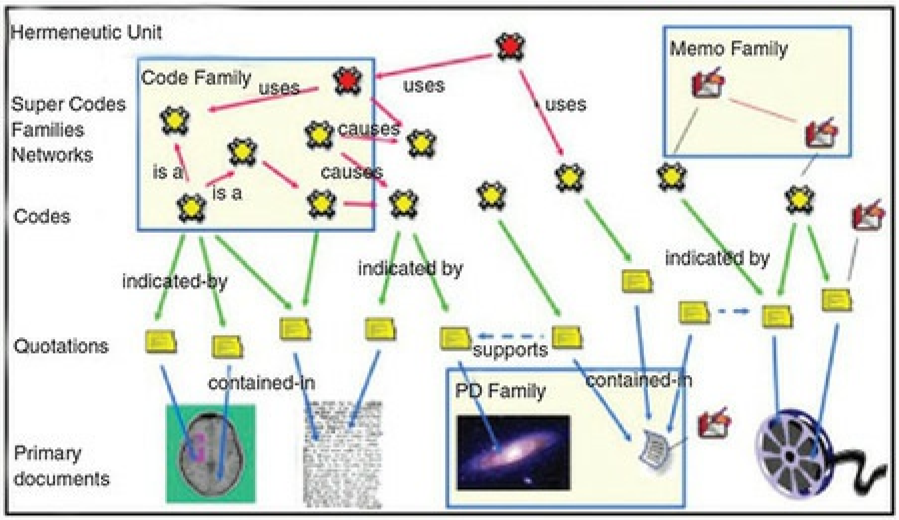
The analysis process can be enhanced in various ways by using a computer. Programs designed for qualitative data can speed up the analysis process, make it easier for researchers to experiment with different codes, test different hypotheses about relationships, and facilitate diagrams of emerging theories and preparation of research reports (Coffey and Atkinson 1996; Richards and Richards 1994). The steps involved in computer-assisted qualitative data analysis parallel those used traditionally to analyze text such as notes, documents, or interview transcripts: preparation, coding, analysis, and reporting. We use three of the most popular programs to illustrate these steps: HyperRESEARCH, QSR NVivo, and ATLAS.ti. (A free trial version of HyperRESEARCH and tutorials can be downloaded from ResearchWare at www.researchware.com.)

Text preparation begins with typing or scanning text in a word processor or, with NVivo, directly into the program's rich text editor. NVivo will create or import a rich text file. HyperRESEARCH requires that your text be saved as a text file (as "ASCII" in most word processors) before you transfer it into the analysis program. HyperRESEARCH expects your text data to be stored in separate files corresponding to each unique case, such as an interview with one subject. These programs now allow multiple types of files, including pictures and videos as well as text. Exhibit 11.20 displays the different file types and how they are connected in the organization of a project (a "hermeneutic unit") with ATLAS.ti.

Computer-assisted qualitative data analysis: The use of special computer software to assist qualitative analyses through creating, applying, and refining categories; tracing linkages between concepts; and making comparisons between cases and events.

Coding the text involves categorizing particular text segments. This is the foundation of much qualitative analysis. Each program allows you to assign a code to any segment of text (in NVivo, you drag through the characters to select them; in HyperRESEARCH, you click on the first and last words to select text). You can make up codes as you go through a document and assign codes that you have already developed to text segments. Exhibit 11.21 shows the screen that appears in HyperRESEARCH and Exhibit 11.22 the screen in NVivo at the coding stage, when a particular text is "autocoded" by identifying a word or phrase that should always receive the same code, or, in NVivo, by coding each section identified by the style of the rich text document—for example, each question or speaker (of course, you should check carefully the results of autocoding). Both programs also let you examine the coded text "in context"—embedded in its place in the original document.

Exhibit 11.20 File Types and Unit Structure in ATLAS.ti


Hermeneutic Unit
Super Codes
Families
Networks
Codes
Quotations
Primary documents
Code Family
uses
uses
uses
causes
causes
is a
is a
Memo Family
indicated-by
indicated by
indicated by
supports
contained-in
contained-in
PD Family

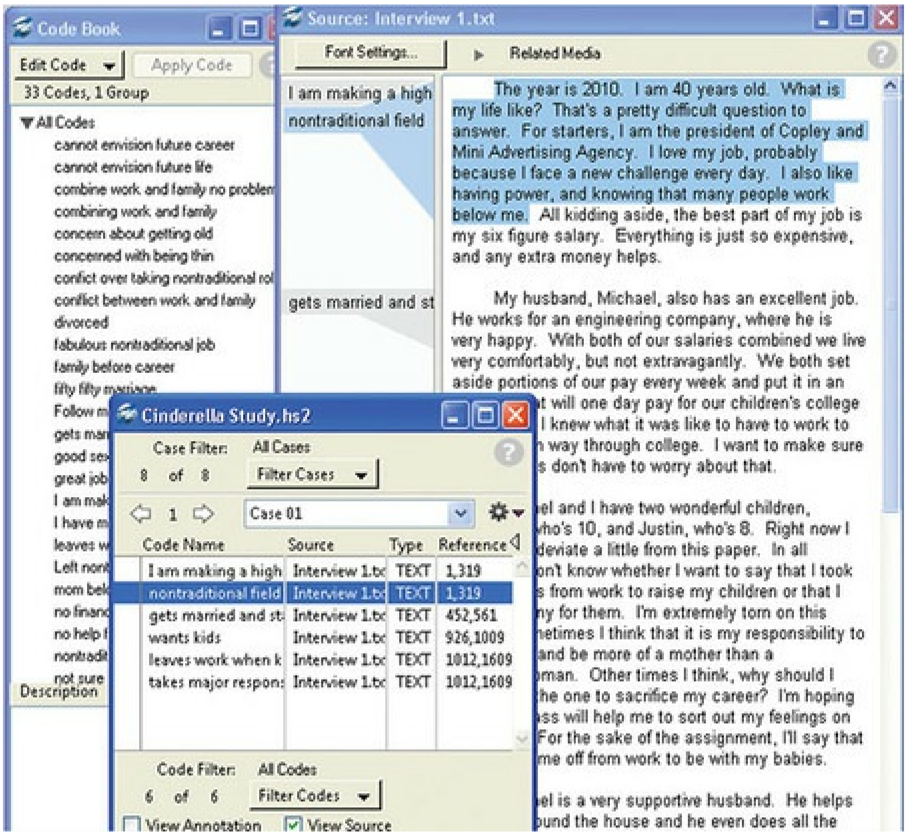
Source: Muhr and Friese (2004:29).

In qualitative data analysis, coding is not a one-time-only or one-code-only procedure. Each program allows you to be inductive and holistic in your coding: You can revise codes as you go along, assign multiple codes to text segments, and link your own comments ("memos") to text segments. You can work "live" with the coded text to alter coding or create new, more subtle categories. You can also place hyperlinks to other documents in the project or to any multimedia files outside it.

Analysis focuses on reviewing cases or text segments with similar codes and examining relationships between different codes. You may decide to combine codes into larger concepts. You may specify additional codes to capture more fully the variation between cases. You can test hypotheses about relationships between codes and develop more free-form models (see Exhibit 11.23). You can specify combinations of codes that identify cases that you want to examine.

Reports from each program can include text to illustrate the cases, codes, and relationships that you specify. You can also generate counts of code frequencies and then import these counts into a statistical program for quantitative analysis. However, the many types of analyses and reports that can be developed with qualitative analysis software do not lessen the need for a careful evaluation of the quality of the data on which conclusions are based.

Exhibit 11.21 HyperRESEARCH Coding Stage

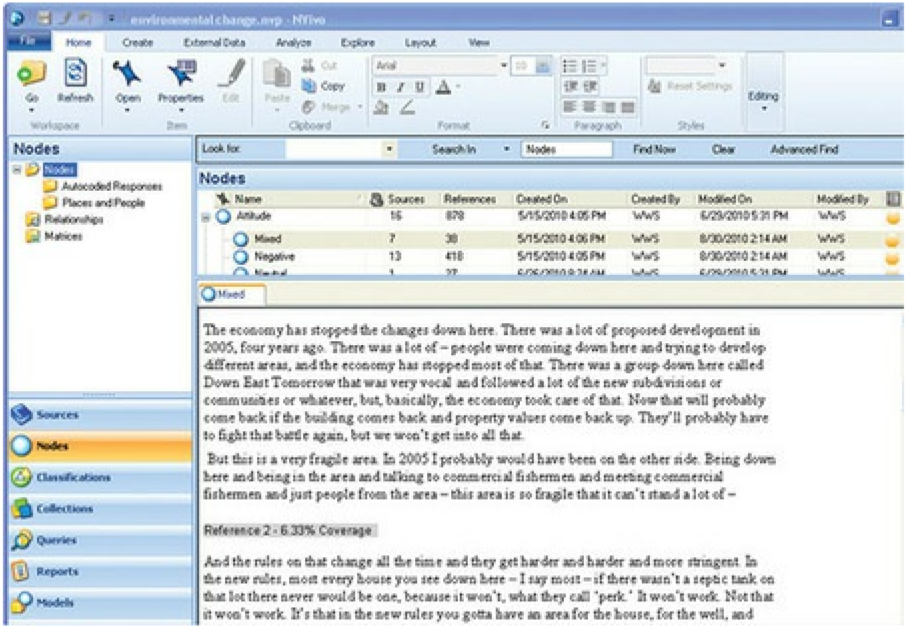

Code Book
Edit Code
Apply Code
33 Codes, 1 Group
All Codes
cannot envision future career
cannot envision future life
combine work and family no problem
combining work and family
concern about getting old
concerned with being thin
conflict over taking nontraditional rol
conflict between work and family
divorced
fabulous nontraditional job
family before career
fifty fifty marriage
Description
Source: Interview 1.txt
Font Settings...
Related Media
I am making a high
nontraditional field
gets married and st
Cinderella Study.hs2
Case Filter: All Cases
8 of 8
Filter Cases
1
Case 01
Code Filter: All Codes
6 of 6
Filter Codes
☐ View Annotation
☑ View Source

The year is 2010. I am 40 years old. What is my life like? That's a pretty difficult question to answer. For starters, I am the president of Copley and Mini Advertising Agency. I love my job, probably because I face a new challenge every day. I also like having power, and knowing that many people work below me. All kidding aside, the best part of my job is my six figure salary. Everything is just so expensive, and any extra money helps.

My husband, Michael, also has an excellent job. He works for an engineering company, where he is very happy. With both of our salaries combined we live very comfortably, but not extravagantly. We both set aside portions of our pay every week and put it in an

Code Name	Source	Type	Reference
I am making a high	Interview 1.tx	TEXT	1,319
nontraditional field	Interview 1.tx	TEXT	1,319
gets married and st	Interview 1.tx	TEXT	452,561
wants kids	Interview 1.tx	TEXT	926,1009
leaves work when k	Interview 1.tx	TEXT	1012,1609
takes major respons	Interview 1.tx	TEXT	1012,1609

Exhibit 11.22 NVivo Coding Stage


environmental change.nvp - NVivo
File
Home
Create
External Data
Analyze
Explore
Layout
View
Go
Refresh
Open
Properties
Edit
Paste
Cut
Copy
Merge
Workspace
Item
Clipboard
Format
Paragraph
Styles
Reset Settings
Editing
Nodes
Autocoded Responses
Places and People
Relationships
Matrices
Look for:
Search In
Nodes
Find Now
Clear
Advanced Find
Sources
Nodes
Classifications
Collections
Queries
Reports
Models
Mixed

Name	Sources	References	Created On	Created By	Modified On	Modified By
Attitude	16	878	5/15/2010 4:05 PM	WWS	6/29/2010 5:31 PM	WWS
Mixed	7	38	5/15/2010 4:06 PM	WWS	8/30/2010 2:14 AM	WWS
Negative	13	418	5/15/2010 4:05 PM	WWS	8/30/2010 2:14 AM	WWS

The economy has stopped the changes down here. There was a lot of proposed development in 2005, four years ago. There was a lot of – people were coming down here and trying to develop different areas, and the economy has stopped most of that. There was a group down here called Down East Tomorrow that was very vocal and followed a lot of the new subdivisions or communities or whatever, but, basically, the economy took care of that. Now that will probably come back if the building comes back and property values come back up. They'll probably have to fight that battle again, but we won't get into all that.

But this is a very fragile area. In 2005 I probably would have been on the other side. Being down here and being in the area and talking to commercial fishermen and meeting commercial fishermen and just people from the area – this area is so fragile that it can't stand a lot of –

Reference 2 - 6.33% Coverage

And the rules on that change all the time and they get harder and harder and more stringent. In the new rules, most every house you see down here – I say most – if there wasn't a septic tank on that lot there never would be one, because it won't, what they call 'perk.' It won't work. Not that it won't work. It's that in the new rules you gotta have an area for the house, for the well, and

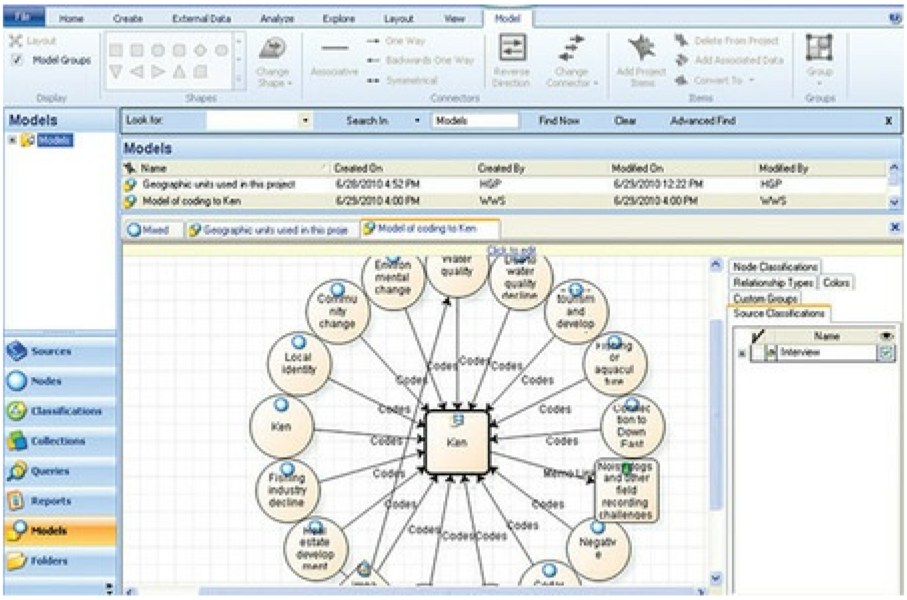
In reality, using a qualitative data analysis computer program is not always as straightforward as it appears. Scott Decker and Barrik Van Winkle (1996) describe the difficulty they faced in using a computer program to identify instances of the concept of *drug sales:*

> The software we used is essentially a text retrieval package. . . . One of the

> dilemmas faced in the use of such software is whether to employ a coding scheme within the interviews or simply to leave them as unmarked text. We chose the first alternative, embedding conceptual tags at the appropriate points in the text. An example illustrates this process. One of the activities we were concerned with was drug sales. Our first chore (after a thorough reading of all the transcripts) was to use the software to “isolate” all of the transcript sections dealing with drug sales. One way to do this would be to search the transcripts for every instance in which the word “drugs” was used. However, such a strategy would have the disadvantages of providing information of too general a character while often missing important statements about drugs. Searching on the word “drugs” would have produced a file including every time the word was used, whether it was in reference to drug sales, drug use, or drug availability, clearly more information than we were interested in. However, such a search would have failed to find all of the slang used to refer to drugs (“boy” for heroin, “Casper” for crack cocaine) as well as the more common descriptions of drugs, especially rock or crack cocaine. (pp. 53–54)

Decker and Van Winkle (1996) solved this problem by parenthetically inserting conceptual tags in the text whenever talk of drug sales was found. This process allowed them to examine all the statements made by gang members about a single concept (drug sales). As you can imagine, however, this still left the researchers with many pages of transcript material to analyze.

Exhibit 11.23 A Free-Form Model in NVivo



Models

Name	Created On	Created By	Modified On	Modified By
Geographic units used in this project	6/28/2010 4:52 PM	HGP	6/29/2010 12:22 PM	HGP
Model of coding to Ken	6/29/2010 4:00 PM	WWS	6/29/2010 4:00 PM	WWS

File
Home
Create
External Data
Analyze
Explore
Layout
View
Model
Shapes
Connectors
Items
Groups
Look for
Search In
Models
Find Now
Clear
Advanced Find
Mixed
Geographic units used in this proje
Model of coding to Ken
Click to edit
Models
Sources
Nodes
Classifications
Collections
Queries
Reports
Models
Folders
Ken
Codes
Environmental change
Water quality
Community change
Local identity
Tourism and develop
Fishing or aquaculture
Ken
Fishing industry decline
Negative
Node Classifications
Relationship Types
Colors
Custom Groups
Source Classifications
Name
Interview

The first step in preparing to analyze qualitative data collected online is to consider what the data represent. Are the data textual, verbal, visual, or some mix of those types? Do the communications represent contacts that are one-on-one, one-to-many, or many-to-many?

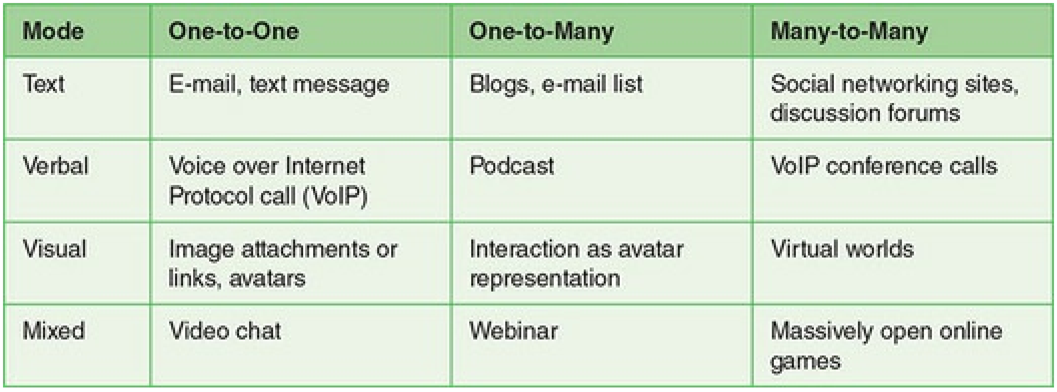
Exhibit 11.24 gives a few examples of each. Your plans also have to take account of whether the material you will study existed without any researcher action (extant), were elicited in response to questions posted by you or another researcher (elicited), or were generated through an interactive process that might have involved reactions to simulations or vignettes (enacted) (Salmons 2016:32). Now that you have defined the type of data you have to work with, you can apply the techniques of qualitative data analysis you learned earlier in this chapter.

Ethics in Qualitative Data Analysis

The qualitative data analyst is never far from ethical issues and dilemmas. Data collection should not begin unless the researcher has a plan that others see as likely to produce useful knowledge. Relations developed with research participants and other stakeholders to facilitate data collection should also be used to keep these groups informed about research progress and findings. Research participants should be encouraged to speak out about emerging study findings (Lincoln 2009:154–155). Decisions to reproduce photos and other visual materials must be considered in light of privacy and copyright issues. Throughout the analytic process, the analyst must consider how the findings will be used and how participants in the setting will react. The need to minimize harm requires attention even after data collection has concluded.

As I noted in Chapter 10, some indigenous peoples have established rules for outside researchers to minimize harm to their community and preserve their autonomy. These rules may require collaboration with an indigenous researcher, collective approval of admission to their culture, and review of all research products before publication. The primary ethical commitment is to the welfare of the community as a whole and the preservation of their culture, rather than to the rights of individuals (Lincoln 2009:162–163).

Exhibit 11.24 Internet Communication Options



Mode	One-to-One	One-to-Many	Many-to-Many
Text	E-mail, text message	Blogs, e-mail list	Social networking sites, discussion forums
Verbal	Voice over Internet Protocol call (VoIP)	Podcast	VoIP conference calls
Visual	Image attachments or links, avatars	Interaction as avatar representation	Virtual worlds
Mixed	Video chat	Webinar	Massively open online games

Source: Salmons, Janet E. 2016. Doing Qualitative Research Online. Thousand Oaks, CA: SAGE.

Participatory action research projects create special challenges because the researcher is likely to feel divided loyalties to the community and the university. Even though it may be difficult for researchers to disagree with the interpretations of community members about the implications of research findings, they may feel that a different interpretation is warranted (Sieber and Tolich 2013:94, 102–104). Qualitative researchers need to be sensitive to the potential for these problems and respond flexibly and respectfully to the concerns of research participants (Hacker 2013:109–110). Allowing community leaders and members to comment on the researcher's interpretations is one way to lessen feelings of exclusion. However, it is also important to recognize that "the community" does not refer to a homogeneous singular entity. As Martin Levinson (2010) explains, "When dealing with human communities[,] groups are actually made up of individuals, and the interests of the group do not necessarily converge with those of the individuals that constitute it" (p. 205).

Qualitative researchers should also take care to consider harm that might occur along all points in the research process. In particular, staff who transcribe interview or focus group transcripts may not seem a part of the formal research process—and may even be independent contractors who are only contacted through the Internet—but they may react emotionally to some of the material in particularly charged interviews. In projects where transcripts contain emotionally charged material, those used as transcriptionists should be given advance warning, and providing connections to counselors should be considered (Sieber and Tolich 2013:175).

Miles and Huberman (1994:293–295) suggest specific questions that are particularly important during the process of data analysis. The next several sections describe these questions.

Privacy, confidentiality, and anonymity.

"In what ways will the study intrude, come closer to people than they want? How will information be guarded? How identifiable are the individuals and organizations studied?" We have considered this issue already in the context of qualitative data collection, but it also must be a concern during the process of analysis. It can be difficult to present a rich description in a case study while not identifying the setting. It can be easy for participants in the study to identify each other in a qualitative description, even if outsiders cannot.

I confronted the challenge of maintaining confidentiality in my analysis of ethnographic notes collected in the study of housing for homeless persons diagnosed with severe mental illness that I used in Chapter 10 to illustrate field notes. I explained the dilemma and the approach I took in the methodological appendix to *Homelessness, Housing, and Mental Illness* (Schutt 2011b):

> It is not possible to understand what happened in the [six] houses, or to consider possible explanations for particular events, without reading detailed descriptions of what people did and having at least some examples of what people said. But providing such details about specific identified group homes would often reveal who it was who was speaking or acting, at least to former housing staff and participants.

> In order to avoid providing intimate and possibly identifiable stories about participants, I have adhered throughout this book to three more specific guidelines. First, I have used arbitrary pseudonyms to refer to everyone whose name appears in the ethnographic notes that I include. Sometimes, I simply refer to individuals by their role, such as "tenant" or "house staff." Second, I have not maintained consistent pseudonyms for individual participants across chapters or even across different incidents that I describe within chapters. This greatly reduces the risk that even a former participant could determine another's identity by gradually being able to imagine the person as I describe different things they said or did. Third, I do not link the activities or statements I report to the individual houses. Since there were only seven different project group homes (six at any one time), it would be much easier for a former participant to link some actions or statements to another participant if the house in which they occurred were identified. However, every group home in our project experienced many of the same types of incidents and interchanges, so by not distinguishing the houses, I left the statements difficult to connect to particular individuals.
>
> Of course this procedure for maintaining participant confidentiality has a cost for the analysis: I cannot "tell the story" of the development of particular houses with the ethnographic details about the houses. However, review of the notes from all the group homes indicates that the similarities in the issues they confronted far outweighed their differences. Moreover, our three ethnographers varied in their styles of observing, interviewing, and even note-taking, so that explicit comparisons of particular houses could be only partial. So maintaining the confidentiality of the individual group homes seems prudent from a methodological standpoint as well as essential from the standpoint of protecting the confidentiality of our research participants. (pp. 317–318)

In ethnographic and other participant studies where initial access is negotiated with community leaders or with groups of participants, qualitative researchers should discuss with participants the approach that will be taken to protect privacy and maintain confidentiality. Selected participants should also be asked to review reports or other products before their public release to gauge the extent to which they feel privacy has been appropriately preserved. Research with photographs that identify individuals raises special ethical concerns. Although legal standards are evolving, it is important not to violate an individual's expectations of privacy in any setting and to seek informed consent for the use of images when privacy is expected (Tinkler 2013:196–198).

Analysis of data collected online creates additional challenges. Guidelines for obtaining voluntary consent were presented in Chapter 10 for data elicited from online participants. When extant data are analyzed, the terms of membership or registration, if any, should be reviewed for the website(s) studied. If it seems that members or participants have been told that their remarks are confidential, or that access to the site is restricted, further efforts are needed to secure participants' consent for use of their statements. It is also a good idea to check with the site moderator, if any, and to review the participants' interactions and statements for evidence of expectations of privacy. If in doubt, posting a statement about the research and requesting comment on the intended use of the data is good practice (Salmons 2016:86–91).

Intervention and advocacy.

"What do I do when I see harmful, illegal, or wrongful behavior on the part of others during a study? Should I speak for anyone's interests besides my own? If so, whose interests do I advocate?" Maintaining what is called *guilty knowledge* may force the researcher to suppress some parts of the analysis so as not to disclose the wrongful behavior, but presenting "what really happened" in a report may prevent ongoing access and violate understandings with participants. The need for intervention and advocacy is more likely to be anticipated in PAR/CBPR projects because they involve ongoing engagement with community partners who are likely to have an action orientation (Hacker 2013:101–104).

Research integrity and quality.

"Is my study being conducted carefully, thoughtfully, and correctly in terms of some reasonable set of standards?" Real analyses have real consequences, so you owe it to yourself and those you study to adhere strictly to the analysis methods that you believe will produce authentic, valid conclusions. Visual images that demean individuals or groups should not be included in publications (Tinkler 2013:197).

Ownership of data and conclusions.

"Who owns my field notes and analyses: I, my organization, my funders? And once my reports are written, who controls their diffusion?" Of course, these concerns arise in any social research project, but the intimate involvement of the qualitative researcher with participants in the setting studied makes conflicts of interest between different stakeholders much more difficult to resolve. Working through the issues as they arise is essential. Mitch Duneier (1999:319–330) decided to end *Sidewalk,* his ethnography of New York City sidewalk book vendors, with an afterword by one of his key informants. Such approaches that allow participants access to conclusions in advance and the privilege to comment on them should be considered in relation to qualitative projects. The public availability of visual images on websites does not eliminate concerns about ownership. Copyright law in the United States as well as in the United Kingdom and Australia provides copyright to content on the Internet as soon as it is uploaded, but there are disagreements about the requirement of informed consent before reproducing images from publicly accessible sites (Tinkler 2013:204–205). Researchers leading PAR/CBPR projects must work out data ownership agreements in advance of data collection to ensure there are no misunderstandings about retention of data and maintenance of confidentiality after the project ends (Hacker 2013:99–100).

Use and misuse of results.

"Do I have an obligation to help my findings be used appropriately? What if they are used harmfully or wrongly?" It is prudent to develop understandings early in the project with all major stakeholders that specify what actions will be taken to encourage appropriate use of project results and to respond to what is considered misuse of these results. Visual researchers must also consider how participants will feel about their images appearing in publications in the future (Wiles et al. 2012):

> People take part in our research, and they don't think in terms of publications arising years and years later. . . . So I think there are lots of problems, even when you have formally and legally the consent they have signed, because it refers to much earlier . . . she might have changed, it's a few years, she might feel very differently, it might remind her now of something very unpleasant. (Youth researcher, focus group 3, p. 48)

Video ethnographers should always make sure that participants know how their images and statements will be used and then avoid depictions that could cause harm to individuals. Consent should be recorded in signed "release and liability waiver statements," although filmed statements of permission for the filming may be appropriate in unplanned situations (Shrum and Scott 2017:121–122).

PAR/CBPR projects are designed to help solve local problems, but harm might also occur if results are not what were expected or if findings cast some elements of the community in an unfavorable light. These possibilities should be addressed as the analysis progresses and resolved before they are publicized (Hacker 2013:114–117).

Conclusions

The variety of approaches to qualitative data analysis makes it difficult to provide a consistent set of criteria for interpreting their quality. Norman Denzin's (2002:362–363) "interpretive criteria" are a good place to start. Denzin suggests that at the conclusion of their analyses, qualitative data analysts ask the following questions about the materials they have produced. Reviewing several of them will serve as a fitting summary for your understanding of the qualitative analysis process:

- *Do they illuminate the phenomenon as lived experience?* In other words, do the materials bring the setting alive in terms of the people in that setting?
- *Are they based on thickly contextualized materials?* We should expect thick descriptions that encompass the social setting studied.
- *Are they historically and relationally grounded?* There must be a sense of the passage of time between events and the presence of relationships between social actors.
- *Are they processual and interactional?* The researcher must have described the research process and his or her interactions within the setting.
- *Do they engulf what is known about the phenomenon?* This includes situating the analysis in the context of prior research and acknowledging the researcher's own orientation on first starting the investigation.

When an analysis of qualitative data is judged as successful in terms of these criteria, we can conclude that the goal of authenticity has been achieved.

As a research methodologist, you should be ready to use qualitative techniques, evaluate research findings in terms of these criteria, and mix and match specific analysis methods as required by the research problem to be investigated and the setting in which it is to be studied.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Case-oriented understanding	431
Computer-assisted qualitative data analysis	451
Conversation analysis	432
Emic focus	415
Ethnomethodology	437
Etic focus	415
Grounded theory	428
Matrix	423
Narrative analysis	435
Participatory action research (PAR)	449
Photo voice	444
Progressive focusing	416
Qualitative comparative analysis (QCA)	438
Systematic observation	446
Tacit knowledge	425
Video ethnography	446
Visual sociology	441

Highlights

- Qualitative data analysts are guided by an emic focus of representing persons in the setting on their own terms, rather than by an etic focus on the researcher's terms.
- Grounded theory connotes a general explanation that develops in interaction with the data and is continually tested and refined as data collection continues.
- Abductive analysis is a theoretically oriented approach to qualitative research that presumes theoretical orientation prior to data collection and engagement with existing scholarship during data analysis.
- Case-oriented understanding provides a description of how participants viewed their experiences in a setting.
- Conversation analysis is a qualitative method for analyzing the sequential organization and details of ordinary conversation.
- Narrative analysis attempts to understand a life or a series of events as they unfolded, in a meaningful progression.
- Visual sociology focuses attention on the record about social life available in photos, videos, or other pictorial displays.
- Photo voice is a visual method in which research participants take pictures of their everyday surroundings with cameras the researcher distributes, and then meet in a group with the researcher to discuss the pictures' meaning.
- Video ethnography involves the use of audiovisual methods and editing techniques to record, analyze, and present one or more viewable social processes, actions, or events in interpretable segments.
- Systematic observation techniques quantify the observational process to allow more systematic comparison between cases and greater generalizability.
- Participatory action research (PAR) uses an ongoing collaboration with community participants to

define the problem for research, to conduct the research, and to develop research reports.

- Computer-assisted qualitative data analysis involves the use of special computer software to assist qualitative analyses through creating, applying, and refining categories; tracing linkages between concepts; and making comparisons between cases and events.

Discussion Questions

1. List the primary components of qualitative data analysis strategies. Compare and contrast each of these components with those relevant to quantitative data analysis. What are the similarities and differences? What differences do these make?
2. Does qualitative data analysis result in trustworthy results—in findings that achieve the goal of authenticity? Why would anyone question its use? What would you reply to the doubters?
3. Narrative analysis provides the “large picture” of how a life or event has unfolded, whereas conversation analysis focuses on the details of verbal interchange. When is each method most appropriate? How could one method add to another?
4. Ethnography (described in Chapter 10), grounded theory, and case-oriented understanding each refer to aspects of data analysis that are an inherent part of the qualitative approach. What do these approaches have in common? How do they differ? Can you identify elements of these three approaches in this chapter’s examples of ethnomethodology, qualitative comparative analysis, visual sociology, and narrative analysis?

Practice Exercises

1. Attend a sports game as an ethnomethodologist seeking to understand how fans construct their sense of the reality of the game. Write up your analysis and circulate it for criticism.
2. Write a narrative in class about your first date, car, college course, or something else that you and your classmates agree on. Then collect all the narratives and analyze them in a "committee of the whole." Follow the general procedures discussed in the example of narrative analysis in this chapter.
3. Go forth and take pictures! Conduct a photo voice project with your classmates and write up your own review of the group's discussion of your pictures.
4. Review one of the articles on the book's study site, at edge.sagepub.com/schutt9e, that used qualitative methods. Describe the data that were collected, and identify the steps used in the analysis. What type of qualitative data analysis was this? If it is not one of the methods presented in this chapter, describe its similarities to and differences from one of these methods. How confident are you in the conclusions, given the methods of analysis used?
5. Find the "Qualitative Data Analysis" lessons in the "Interactive Exercises" link on the study site. Answer the questions in this lesson and read the corresponding article.

Ethics Questions

1. Pictures are worth a thousand words, so to speak, but is that 1,000 words too many? Should qualitative researchers (like yourself) feel free to take pictures of social interaction or other behaviors anytime, anywhere? What limits should an institutional review board place on researchers' ability to take pictures of others? What if the picture of the empty table in this chapter also included the abusive family member who is discussed? What if the picture was in a public park, rather than in a private residence?
2. Participants in social settings often "forget" that an ethnographer is in their midst, planning to record what they say and do, even when the ethnographer has announced his or her role. New participants may not have heard the announcement, and everyone may simply get used to the ethnographer as if he or she was just "one of us." What efforts should ethnographers take to keep people informed about their work in the settings they are studying? Consider settings such as a sports team, a political group, and a book group.

Web Exercises

1. *Qualitative Research* is an online journal about qualitative research. Inspect the table of contents for a recent issue at http://journals.sagepub.com/home/qrj. Read one of the articles and write a brief article review.
2. Be a qualitative explorer! Go to the “Qualitative Page” website and review some of the qualitative resources available for different “Research Approaches.” (https://qualpage.com/2016/08/01/qualpage-relaunched/). Be careful to avoid textual data overload.

Video Interview Questions

Listen to the researcher interviews for Chapter 11 at edge.sagepub.com/schutt9e.

1. Paul Atkinson believes that researchers should think about not only what people are talking about but also “how” they are talking about a topic or concept. Do you agree with this statement? Why or why not?
2. What are Atkinson’s three suggestions for dealing with narratives?

HyperRESEARCH Exercises

1. Eight essays written by college-age women in Sharlene Hesse-Biber's (1989) "Cinderella Study" are available publicly. The essays touch on how young women feel about combining work and family roles. Download and install HyperRESEARCH from http://www.researchware.com/downloads/downloads-hyperresearch.html and open the Cinderella Study (start the Tutorials first, then double-click on the Cinderella Study folder). Look over the preliminary code categories that have already been applied to each essay. Do you agree with the code categories/themes already selected? What new code categories would you add and why? Which would you delete? Why? What are some of the common themes/codes that cut across all eight cases concerning how young women think about what their life will be like in 20 years?
2. Work through the tutorials on HyperRESEARCH that was downloaded with the program. How does qualitative analysis software seem to facilitate the analysis process? Might it hinder the analysis process in some ways? Explain your answers.

Developing a Research Proposal

1. Identify the qualitative data analysis alternative that is most appropriate for the qualitative data you proposed to collect for your project (Exhibit 3.10, #18).
2. Using the approach selected in #1, develop a strategy for using the techniques of qualitative data analysis to analyze your textual data.

Section IV Complex Social Research Designs

Chapter 12 Mixed Methods

- Research That Matters, Questions That Count
- History of Mixed Methods
- Types of Mixed Methods
  - Integrated Designs
  - Embedded Designs
  - Staged Designs
  - Complex Designs
- Strengths and Limitations of Mixed Methods
- Research in the News: Why Women Don't Report Sexual Harassment
- Careers and Research
- Ethics and Mixed Methods
- Conclusions

Two young teenagers in upstate New York are detained by police for defacing property, "Joe" for painting graffiti on the side of a store and "Sam" for painting graffiti on a downtown sidewalk. After reviewing the cases, a probation officer in the local family court's intake unit refers Joe for adjudication in court and diverts Sam from court so that he can try to resolve his case through counseling and a restitution plan.

You learn these facts about the two cases 5 years after they occurred during your investigation of why some juveniles are treated more harshly than others. Your research design involves analyzing information in the juvenile intake unit's records for a 1-year period. You record age, sex, race, education, and family structure for a sample of juvenile offenders as well as the characteristics of the offense they are charged with and their record of prior offenses. You also record the severity of the disposition of each case: Referral to a court hearing is coded as a "severe" disposition and diversion from court is coded as a "lenient" disposition. Would you wonder why Joe was treated more harshly than Sam?

This is the basic procedure used in Carolyn Needleman's (1981) investigation of a juvenile justice system in an area of New York, and it is the question that led her to add a qualitative component to her quantitative study—to use *mixed methods.* This procedure is similar to those relied on by many others studying juvenile justice, including Dale Dannefer and me (Dannefer and Schutt 1982; Schutt and Dannefer 1988). After coding these data, Needleman used statistics to determine whether juvenile offenders were treated more harshly if they were older, if they were African American, if they were male, if they already had a record, and so on. But Needleman also sought to understand the processing of juvenile cases, and so she observed the probation officers in the intake office and interviewed them repeatedly over a 2-year period.

Research That Matters, Questions That Count

Violence against women on college campuses is often attributed to excessive drinking and for that reason its seriousness is too often minimized. Maria Testa at the University at Buffalo's Research Institute on Addictions decided that the complex contextual and social factors involved in incidents of violence against women had to be investigated so that we can learn why such violence occurs and how alcohol is involved. Although she had been trained as a quantitative researcher, her experiences had led her to the realization that for such research questions it could help to complement quantitative data with qualitative data.

The resulting research project carried out by Maria Testa, Jennifer Livingston, and Carol VanZile-Tamsen used a mixed-methods design to identify women who had been the victims of violence and then develop an in-depth picture of how the violence had occurred. Through analysis of the detailed descriptions that women provided of the incidents when they had been victimized, Testa and her coauthors found that the common assumption that sexual victimization results from women's impaired judgment was not correct. Instead, rape occurred after excessive drinking when women were incapacitated and could neither resist nor often even remain fully aware of what was happening to them. Testa termed these incidents *incapacitated rape.*

Testa and her colleagues concluded that insights yielded by the qualitative data analysis fully justified the time-consuming process of reading and rereading interviews, coding text, and discussing and reinterpreting the codes.

1. What other qualitative methods could be used to study violence against women on college campuses? What would be the advantages and disadvantages of focus groups or observation?
2. If you were asked to design a study of violence against women on a college campus and then develop a comprehensive report about the problem for the student senate, would you put more emphasis on quantitative methods or qualitative methods, or would you attempt to use both equally? Explain your reasoning.

In this chapter, you will learn about the logic and procedures for designing mixed-methods research projects as well as about the research on violence. By the end of the chapter, you will understand why a mixed-methods design can be the most advantageous design for investigating complex social patterns. In addition, you will be able to identify features that distinguish particular mixed-methods designs and evaluate their appropriateness for particular research questions. After you finish the chapter, test yourself by reading Testa et al.'s 2011 article in the journal *Violence Against Women* at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 12 at edge.sagepub.com/schutt9e.

Source: Testa, Maria, Jennifer A. Livingston, and Carol VanZile-Tamsen. 2011. "Advancing the Study of Violence Against Women Using Mixed Methods: Integrating Qualitative Methods Into a Quantitative Research Program." *Violence Against Women* 17(2):236–250.

Like the conclusion by Maria Testa and her colleagues that impaired judgment was not the immediate cause of campus sexual assault, what Carolyn Needleman learned from her qualitative participant observation surprised her: Probation workers did not regard a referral to court as a harsher sanction than being diverted from court (Needleman 1981:248). In fact, the intake workers believed the court was too lenient in its decisions and instead felt their own "adjustment" effort during probation, with counseling, negotiation, restitution, and social service agencies was harsher and more effective (Needleman 1981:249):

> Considered toothless, useless and counterproductive, the court represents to the intake workers not a harsher sanction, but a dumping ground for the "failures" from the intake's alternative [probation]. (Needleman 1981:253)

In this way, both Needleman and Testa found that concepts measured with official records or as believed by the public differ markedly from the meaning attached to these records by probation officers. Needleman's qualitative interviews revealed that probation officers often diverted cases from court because they thought the court would be too lenient, without learning anything about the individual juvenile. Perhaps most troubling for research using case records, Needleman found that probation officers decided how to handle cases first and then created an official record that appeared to justify their decisions. Exhibit 12.1 summarizes the differences Needleman's qualitative interviews identified between researchers' assumptions about juvenile court workers and the juvenile court workers' own assumptions.

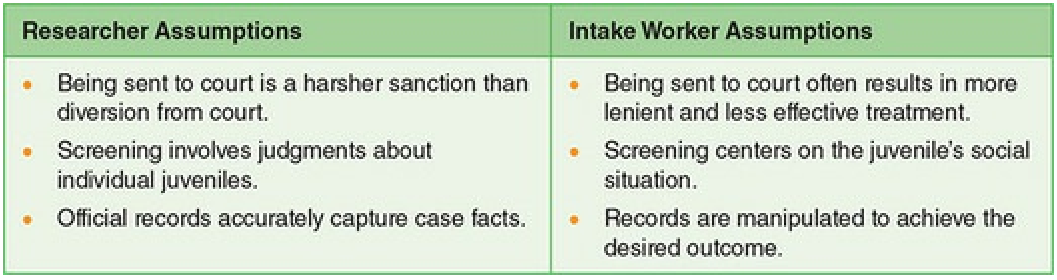
The vignettes from Testa and Needleman illustrate the rationale for mixed methods: A single method may not represent adequately the social world's complexity that we are trying to understand. The goal of this chapter is to highlight the importance of this lesson and to identify its implications for research practice. I will discuss the historical and philosophical foundation for mixed methods, illustrate the different ways that methods can be combined, and highlight some of the advantages and challenges of using multiple methods in one study. Although in other chapters I have mentioned some investigations that combined different methods, what is distinctive about this chapter is its focus on research designs that have major qualitative and quantitative components as well as its attention to the challenges of combining methods.

History of Mixed Methods

Sociologists and other social scientists have long used multiple methods in their research, but only in recent decades have some focused attention on how best to combine qualitative and quantitative methods to better achieve research goals. As an example of the classic tradition, anthropologist W. Lloyd Warner, Marchia Meeker, and Kenneth Eells's (1960) classic *Social Class in America* used both qualitative and quantitative research methods to understand the social complexities of U.S. communities. Based largely on his pioneering field research in the 1930s and 1940s in the New England community of Newburyport—called "Yankee City" in his subsequent books—Warner showed how qualitative data about social position and activities could be converted into a quantitative measure that he termed the "Index of Status Characteristics" (I.S.C.). Here is an example of how Warner et al. (1960) shifted from qualitative description to quantitative assessment:

> The home of Mr. and Mrs. Henry Adams Breckenridge . . . has three stories and is topped by a captain's walk. . . . Large trees and a tall thick hedge . . . garden stretches one hundred yards . . . many old rose bushes. . . . The life and surroundings of Mrs. Henry Adams Breckenridge, old-family and upper-upper, . . . Her ratings of the characteristics of her I.S.C. give her a final score of 12, or perfect. . . . (pp. 243, 247)

Exhibit 12.1 Researchers' and Juvenile Court Workers' Discrepant Assumptions



Researcher Assumptions	**Intake Worker Assumptions**
• Being sent to court is a harsher sanction than diversion from court. • Screening involves judgments about individual juveniles. • Official records accurately capture case facts.	• Being sent to court often results in more lenient and less effective treatment. • Screening centers on the juvenile's social situation. • Records are manipulated to achieve the desired outcome.

Source: Needleman (1981:248–256).

Another classic example of the early use of mixed methods is represented in *Union Democracy,* sociologist Seymour Martin Lipset's study of the International Typographical Union (ITU) with Martin Trow and James Coleman (1956). After some preliminary investigations involving reading union literature and interviewing ITU members, Lipset began systematic qualitative research.

> Aside from long exploratory interviews with key informants in the union,

> members of the research team that was organized around the study familiarized themselves in every way possible with the actual political life of the union, attending union meetings, party caucuses, and chapel meetings, while paying particular attention to the events preceding the local union election held in May 1951. (Lipset et al. 1956:xiii).

The qualitative research "led to a sharpening and respecifying of hypotheses" and the decision to add quantitative methods.

> At this point it seemed that certain crucial aspects of the internal political process of the ITU could best be studied through survey research methods, and moreover, that such a study could be practically carried out among the members of the New York local. (Lipset et al. 1956:xiii–xiv)

Throughout *Union Democracy,* Lipset et al. (1956:470–492) analyzed both qualitative and quantitative data to identify and illustrate factors that influence political practices in the union. However, they did not discuss the ways in which their use of both quantitative and qualitative methods helped them to answer research questions or created particular challenges.

Donald Campbell and Donald Fiske (1959) were an early exception to this history of avoiding explicit discussion of mixed methods. They proposed that a "multitrait–multimethod matrix" in validity studies would allow greater confidence in the validity of measures when the results of measuring the same phenomenon with different methods converge—and when the results of measuring different phenomena with the same method diverge. This article quickly became a classic, but there was little additional explicit attention to the value of mixed methods for the next two decades.

Mixed methods: Research that combines qualitative and quantitative methods in an investigation of the same or related research question(s).

Multitrait–multimethod matrix: A method of evaluating the validity of measures by determining whether measuring the same phenomenon with different methods leads to convergent results and measuring different phenomena with the same method leads to divergent results.

Not until 1989 did John Brewer and Albert Hunter's book *Multi-Method Research* herald more systematic attention to this approach. At that time, the use of multiple methods to study one research question was often called triangulation (see pp. 124–125). The term suggests that a researcher can get a clearer picture of the social reality being studied by viewing it from several different perspectives. Each will have some liabilities in a specific research application, and all can benefit from a combination of one or more other methods (Brewer and Hunter 1989; Sechrest and Sidani 1995).

Increasing interest in mixed methods reflected to some extent the increasingly complex research problems that social researchers were confronting and the demands of policy makers and others for sophisticated, convincing evidence. Mixed methods have been used more in applied fields than in academic disciplines: 16% of articles in nursing and education used mixed methods in a sample drawn in 2005, compared with 6% of articles in sociology and psychology (Alise and Teddlie 2010:115). There is also evidence in many publications of growing acceptance of the value of qualitative methods and of recognition by many qualitative researchers of the need for more generalizable quantitative evidence (Alise and Teddlie 2010:115; Creswell and Plano Clark 2011:20–22). But the turn toward mixed methods in the late 1980s also marked the decline in disagreements between qualitative and quantitative methodologists about the philosophies behind their methodological preferences. After this period of conflict known as the "paradigm wars," social researchers were more willing to consider qualitative and quantitative methods as complementary (Alise and Teddlie 2010).

Marcus Weaver-Hightower (2013) explained the basis for his rejection of the paradigm wars approach in his report on a mixed-methods study of policy making about boys' education in Australia:

> Both qualitative and quantitative methods do have limitations for studying policy influence; qualitative methods can have difficulty establishing the extent of influence while quantitative methods can have difficulty providing the whys, hows, and so whats. Rather than succumb to paralysis from competing claims for methodological incompleteness, I used mixed methods to ameliorate each approach's limitations. (p. 6)

Ryan Brown and his colleagues (2013) at the RAND Corporation were also guided by this logic when they sought to learn why most health intervention strategies with homeless adults have had limited effect:

> Properly addressing such complexity and interdependency requires agile research strategies that can not only assess causal factors at multiple levels but also flexibly incorporate new information as it arises during the research process. This means enabling creative and productive conversation between qualitative and quantitative measurement and analytic modalities—a mixed-methods approach. (p. 329)

Paradigm wars: The intense debate from the 1970s to the 1990s between social scientists over the value of positivist and interpretivist/constructivist research philosophies; also see *scientific paradigm*.

Types of Mixed Methods

A researcher who plans to use both qualitative and quantitative research methods to best address a research problem is adopting a mixed-methods approach (Plano Clark and Ivankova 2016:57). However, the specific ways to "mix" the methods vary between research problems and settings. Qualitative methods may be used before or after quantitative methods; that is, the sequencing of the two approaches can differ. In addition, one method may be given priority in a project, or qualitative and quantitative methods may be given equal priority. Distinguishing the sequencing and priority of the qualitative and quantitative methods used in any mixed-methods project results in a number of different types. Although some mixed-methods researchers use different labels for the types, and others believe different methods should just be combined as needed in a project without worrying about "types," I think you'll find that reading about the types will give you a better sense of the options (Plano Clark and Ivankova 2016:126–128).

Before discussing these types, it will help to learn some basic conventions for naming them:

- The primary method used in a mixed-methods project is written in all caps (QUAN or QUAL).
- The secondary method is written in lowercase letters (quan or qual).
- If both methods are given equal priority, they are both written in all caps.
- If one method is used before the other, the sequence is indicated with an arrow (QUAL→quan, or Qual→QUAN, or QUAN→QUAL, etc.).
- If two methods are used concurrently, but one has priority, the secondary method is said to be "embedded" in the primary method. This is indicated as follows: QUAL(quan) or QUAN(qual).
- If two methods are used concurrently, but they have equal priority, the relation between the two methods is indicated with a + symbol: QUAL+QUAN.

The different types of mixed methods that result from these distinctions in priority and sequence are represented in Exhibit 12.2. Of course, these different types represent only research projects in which each method is used only once; more complex mixed-methods projects may involve combinations of these types, either at the same time or in sequence.

Some examples using the research projects I have already introduced will help to clarify the reasons for using these different types (Creswell and Plano Clark 2011:8; Morgan 2014:20). Needleman (1981) collected quantitative data from case records on dispositions in juvenile cases, but she realized that she might misinterpret the meaning of these numbers unless she also understood the meaning behind the probation officers' dispositions. She therefore embedded a qualitative study of probation officers within her larger quantitative study of case dispositions. This project can be represented as using a QUAN(qual) embedded method.

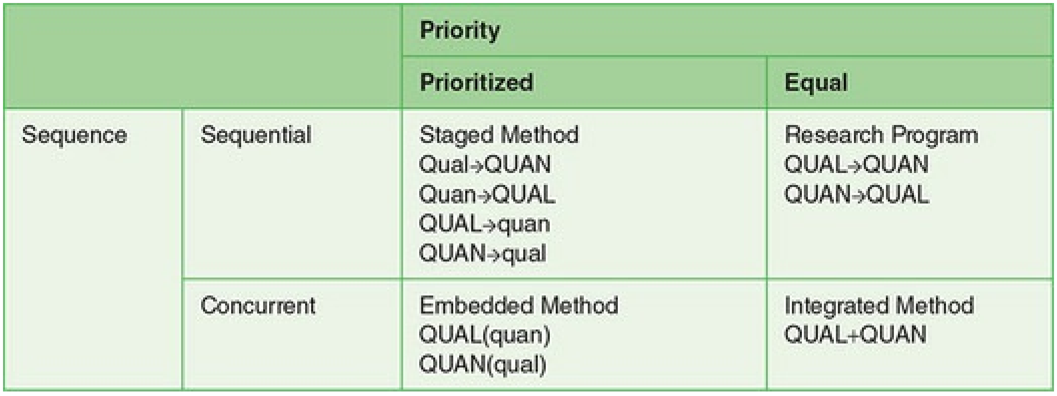
In their study of the International Typographical Union, Lipset and colleagues (1956) needed to identify the differences in orientations of the union's various social groups and to understand how these differences were expressed in union politics. For this purpose, they used an integrated method (QUAL+QUAN) in which both quantitative methods (a structured survey) and qualitative methods (in-depth interviews) played major roles.

You learned in the section on refining and testing questions in Chapter 8 that qualitative methods like focus groups and cognitive interviews can be used to refine questions for a structured quantitative survey: qual→QUAN. In the section on organizing concepts in Chapter 10, you saw that observations may be coded to create systematic quantitative data: QUAL→quan. In these types of staged methods, one method has priority. In addition, the use of one method precedes the other one in time because the data collected with one method shape the data collected with the other method.

The various studies of the response to domestic violence that I described in Chapter 2 represent a common way of combining methods in an extended research program with mixed methods, in which the results of research using one method inform or raise questions about research focused on the same research question that used a different research method. For example, Angela Moe's research (2007) on victim orientation to calling the police after an experience of abuse (Chapter 2's section on exploratory analysis) represented another step in the research program that began with Lawrence Sherman and Richard Berk's (1984) quantitative study of the police response to domestic violence. This would be a QUAN→QUAL research program.

Research program with mixed methods: Qualitative and quantitative methods are used in sequence and are given equal priority.

Exhibit 12.2 Types of Mixed Methods



		Priority	
		Prioritized	**Equal**
Sequence	Sequential	Staged Method Qual→QUAN Quan→QUAL QUAL→quan QUAN→qual	Research Program QUAL→QUAN QUAN→QUAL
	Concurrent	Embedded Method QUAL(quan) QUAN(qual)	Integrated Method QUAL+QUAN

In this section, I will discuss in more depth examples of research projects using integrated, staged, and embedded methods. When qualitative and quantitative methods are combined in these ways, researchers must make key decisions about how best to mix methods to achieve their study purposes. As you see how researchers make these decisions, you will also understand better the strengths and weaknesses of the specific methods.

Integrated Mixed-Methods Designs

In an integrated mixed-methods design, qualitative and quantitative methods are used concurrently and both are given equal importance. Findings produced from these methods are then integrated and compared during the analysis of project data. This is the QUAL+QUAN design. Susan McCarter (2009) extended prior research on juvenile justice processing with an integrated mixed-methods investigation of case processing and participant orientations in Virginia. Her hope was to use the results of the two methods to triangulate the study's findings; that is, to show that different methods lead to similar conclusions and therefore become more credible.

The large quantitative data set McCarter (2009) used in her research was secondary data collected on 2,233 African American and Caucasian males in Virginia's juvenile justice system:

> The quantitative data set (n = 2,920) is a disproportionate, stratified, random sample of juvenile cases from all 35 Virginia Court Service Units (CSU) where each CSU was treated as a separate stratum. These data were collected by the Joint Legislative Audit and Review Commission (JLARC) in an examination of court processing and outcomes of delinquents and status offenders in Virginia.
>
> JLARC collected data on the juveniles' previous felonies; previous misdemeanors; previous violations of probation/parole; previous status offenses; recent criminal charges, intake action on those charges, pre-disposition(s) of those charges, court disposition(s) of those charges; and demographics such as sex, race, data of birth, CSU, and geotype (urban, suburban, rural). For a subset of these cases, data included information from the youth's social history, which required judicial request. (p. 535*)

Qualitative data were obtained from 24 in-depth interviews with juvenile judges, the commonwealth's attorneys, defense attorneys, police officers, juveniles, and their families (McCarter 2009):

> The juvenile justice personnel were from six Court Service Units across the state, including two urban, two suburban, two rural, two from Region I, two from Region II, and two from Region III. . . . Participants from each CSU were chosen to provide maximum diversity in perspectives and experiences, and thus varied by race, sex, and age; and the justice personnel also varied in length of employment, educational discipline and educational attainment. (p. 536) . . .

> The youth and their families were all selected from one Court Service Unit (CSU) located in an urban geotype with a population of approximately 250,000 (p. 536). The sample of youth and their family members was comprised of all male juveniles, five mothers and one father. Four of the six families were African American and two were Caucasian. (p. 540[*])

[*] McCarter, Susan A. 2009. "Legal and Extralegal Factors Affecting Minority Overrepresentation in Virginia's Juvenile Justice System: A Mixed-Method Study." *Child and Adolescent Social Work Journal* 26:533–544. Copyright © 2009, Springer. Reprinted with permission.

Integrated mixed-methods design: Qualitative and quantitative methods are used concurrently and both are given equal importance.

The in-depth interviews included both open- and closed-ended questions. The open-ended responses are coded into categories that distinguished how participants perceived the role of race in the juvenile justice system (McCarter 2009:536). A direct connection with the quantitative findings was made in the interviews themselves:

> Respondents were read the quantitative findings from this study and then asked whether or not their experiences and/or perceptions of the juvenile justice system were congruent with the findings. They were also asked how commonly they believed instances of racial or ethnic bias occurred in Virginia. (McCarter 2009:540)

Comments made in response to this qualitative question supported the quantitative finding that race mattered in the juvenile justice system:

> Juvenile justice professionals as well as youth and their families cited racial bias by individual decision-makers and by the overall system, and noted that this bias was most likely to occur by the police during the Alleged Act or Informal Handling stages. However, although race was considered a factor, when compared to other factors, professionals did not think race played a dominant role in affecting a youth's treatment within the juvenile justice system. . . . Eighteen of the juvenile justice professionals stated that they felt a disparity [between processing of African American and white juveniles] existed, four did not feel that a disparity existed, and two indicated that they did not know. (McCarter 2009:540)

In this way, the qualitative and quantitative findings were integrated and the study's key conclusion about race-based treatment was strengthened because it was based on triangulated identification (McCarter 2009:542)—supporting earlier conclusions of quantitative research by Dannefer and Schutt (1982).

Embedded Mixed-Methods Designs

Testa and colleagues (2011) supplemented their quantitative study of violence against women with a qualitative component because violence against women is “a complex, multifaceted phenomenon, occurring within a social context that is influenced by gender norms, interpersonal relationships, and sexual scripts” and “understanding of these experiences of violence is dependent on the subjective meaning for the woman and cannot easily be reduced to a checklist” (p. 237). This was an embedded mixed-methods design (QUAN[qual]).

Embedded mixed-methods design: Qualitative and quantitative methods are used concurrently in the research but one is given priority.

Victims’ responses to structured questions indicated an association between alcohol and rape, but when victims elaborated on their experiences in qualitative interviews, their comments led to a new way of understanding this quantitative association. Although this association has often been interpreted as suggesting “impaired judgment” about consent by intoxicated victims, the women interviewed by Testa et al. (2011) all revealed that they had had so much to drink that they were unconscious or at least unable to speak at the time of the rape. Testa and her colleagues (2011) concluded that the prevalence of this type of “incapacitated rape” required a new approach to the problem of violence against women:

> Qualitative analysis of our data has resulted in numerous “a-ha” types of insights that would not have been possible had we relied solely on quantitative data analysis (e.g., identification of incapacitated rape and sexual precedence, heterogeneity in the way that sexual assaults arise) and also helped us to understand puzzling quantitative observations. . . . These insights, in turn, led to testable, quantitative hypotheses that supported our qualitative findings, lending rigor and convergence to the process. We never could have anticipated what these insights would be and that is what is both scary and exhilarating about qualitative data analysis, particularly for a scientist who has relied on quantitative data analysis and a priori hypothesis testing. The lengthy process of reading, coding, rereading, interpreting, discussing, and synthesizing among two or more coders is undeniably a major investment of time. (p. 242)

Testa and her colleagues concluded that insights yielded by the qualitative data analysis fully justified the time-consuming process of reading and rereading interviews, coding text, and discussing and reinterpreting the codes (p. 245).

Staged Mixed-Methods Designs

Migration in search of employment creates family strains and lifestyle changes in countries around the world. This is nowhere as true as in contemporary China, where millions of young people have migrated from rural villages to industrial cities. Juan Zhong and Jeffrey Arnett (2014) sought to understand the impact of such migration on Chinese women workers' views of themselves as adults. The researchers chose for this purpose a staged mixed-methods design (QUAN-qual), in which a quantitative survey of 119 women workers, aged 18–29, from a factory in Guangdong, China, preceded and was the basis for qualitative interviews with 15 of them.

Staged mixed-methods design: Qualitative and quantitative methods are used in sequence in the research and one is given priority.

Zhong and Arnett posed three research questions:

1. To what extent do young Chinese migrant women workers consider themselves to have reached adulthood?
2. What are the conceptions of adulthood held by young Chinese migrant women workers, that is, what criteria are most important?
3. What is the relationship between marital status (or parental status) and conceptions of adulthood?

The researchers explained their use of mixed methods as reflecting the complexity of the phenomena involved with these questions:

> Given the complexity of social phenomena, mixed-methods research allows investigators to examine a problem from different but complementary perspectives, so that they can obtain a multilevel and contextual understanding of the phenomenon (Creswell and Plano Clark 2011). . . . In this study, explanatory design was implemented, in which qualitative findings were used to help interpret and contextualize quantitative results. (p. 257)

The structured questionnaire included questions about feelings of adulthood and the importance attached to various markers as indicating whether a person was an adult. Only 44% responded that they felt like they had reached adulthood, whereas most of the rest indicated that they felt adult in some ways but not in others. The five markers rated most often as important were (1) Learn to care for parents, (2) If a man, become capable of supporting a family financially, (3) Settle into a long-term career, (4) If a woman, become capable of caring for children, and (5) If a man, become capable of keeping family physically safe.

Analysis of the survey data indicated that being married and having children were associated with feeling like an adult, irrespective of age. The three markers of adulthood identified as most important in the quantitative survey were “Learn to care for parents,” “Settle into a long-term career,” and “Become capable of caring for children.”

The qualitative component began with the random selection of 15 of the women workers who had responded to the questionnaire.

> Two of the questions were asked to tap into their conceptions of adulthood. One was ‘‘Do you feel like you have reached adulthood? Why or why not?’’ which aimed to find out how the women workers viewed their transition to adulthood, and how they applied the markers of adulthood to their own life. The other question was about their more general conceptions of the transition to adulthood: “What do you think are the most important markers of adulthood?” Compared with the brief choice on the questionnaire, this question in the interview allowed the participants to describe their understanding of transition to adulthood in more detail and further explain which markers they thought were important for being an adult and why. (Zhong and Arnett 2014:258)

The qualitative analysis began with Zhong reading the transcripts and coding the themes identified about conceptions of adulthood. Arnett replicated this process, and the two compared their results and resolved discrepancies. Ultimately, they settled on four major domains that captured the expressed views about adulthood: family obligations and capacities, relational maturity, role transitions, and individualism. In the following sections of their article, each domain of markers for adulthood was elaborated with illustrations from the interviews.

As in the quantitative survey, the theme of family obligations and capacities was often emphasized as associated with feelings of adulthood. A quote from a married 23-year-old woman who had a 1-year-old son provided more insight into the basis for this association:

> I am married now, I should have a sense of responsibility toward family; I am also a mother, I should take care of my child. My parents-in-law are taking care of him, and we must send a few thousand yuan to them every year. Right now, he is only 1 year old. When it’s time for him to go to school, I will go back home and take care of him myself. After all, my parents-in-law, who are not well educated, only care about if he is full [well-fed] or warm. They don’t know how

to teach him, so it’s better for me to take care of him. (Zhong and Arnett 2014:260)

Complex Mixed-Methods Designs

Brown and his colleagues (2013) used a mixed-methods design in their study of sexual intimacy among homeless men in downtown Los Angeles. This study can best be understood as using a complex mixed-methods design because it involved a clear sequence in data collection as well as an integrated analysis in which both QUAL and QUAN methods had equal importance. The researchers first used qualitative interviews and then a structured quantitative survey, without giving either method priority. After they had collected data with both these QUAL and QUAN methods, Brown et al. developed an integrated analysis of the findings. The overall design might be diagrammed as a (QUAL→QUAN)→(QUAL+QUAN) design.

> Complex mixed-methods design: Qualitative and quantitative methods are combined in a research project that uses more than one of the four basic types of mixed-methods designs or that repeats at least one of those basic types.

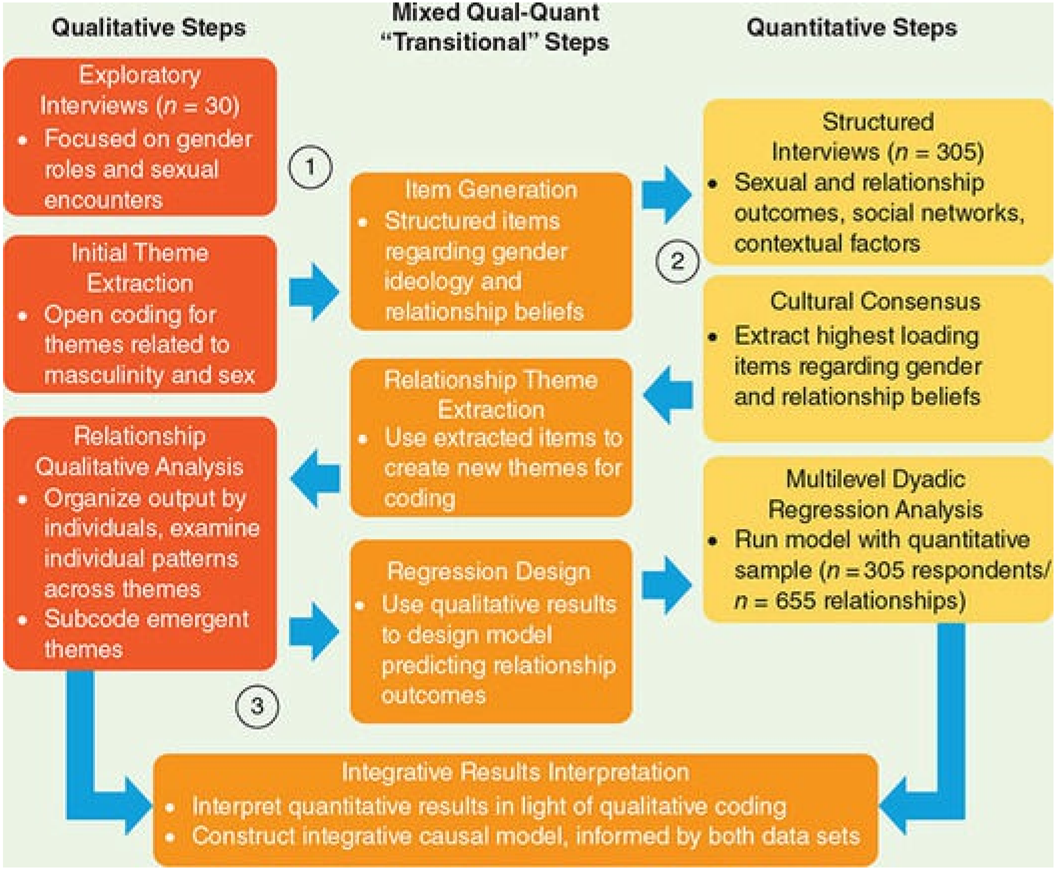
The qualitative research began with a sample of 30 persons using two meal lines and three shelters in the Los Angeles Skid Row area. Thirty of 36 men identified as eligible completed audiotaped interviews of 1 to 1 1/2 hours, for which they received $25. The researchers then finalized a structured questionnaire, based partly on what they learned in the qualitative interviews, and distributed this questionnaire several months later to a random sample of 338 men selected from 13 meal lines identified in service directories and provider interviews; 305 men completed the survey after giving their consent and so received a $25 incentive. The issues addressed in the qualitative interviews and structured survey overlapped in coverage of gender ideologies and sexual behaviors, and the structured survey included questions about individual characteristics, sexual partners, and social network characteristics (Brown et al. 2013:329–330).

Exhibit 12.3 represents the different elements and stages of Brown et al.'s collection and analysis of data. In the first exploratory step, they conducted qualitative interviews and extracted themes from the resulting transcripts. These themes were then used in the second step to generate questions for their structured interviews. After they conducted the interviews, analysis of the structured data identified beliefs that were associated with each other and so suggested themes that could be examined with the qualitative data. Further analysis of the qualitative data then suggested more analyses with the quantitative data. Finally, the results of the qualitative and quantitative analyses were integrated into a combined explanation for the men's behavior.

The analysis made possible by this integrated use of qualitative and quantitative data was actually even more complex than the exhibit indicates. In the authors' words:

> With interlaced qualitative and quantitative data collection and analysis, our initial gender-based model was subject to multiple revisions and modifications over time. Multiple rounds of qualitative and quantitative analysis (including multiple iterations of hypothesis generation and testing) moved us away from thinking about gender ideology and hypermasculinity and toward thinking about gender roles and structural barriers to enacting these roles. (Brown et al. 2013:330)

Exhibit 12.3 Qualitative and Quantitative Methodological Steps


Qualitative Steps
Mixed Qual-Quant "Transitional" Steps
Quantitative Steps
Exploratory Interviews (n = 30)
• Focused on gender roles and sexual encounters
Initial Theme Extraction
• Open coding for themes related to masculinity and sex
Relationship Qualitative Analysis
• Organize output by individuals, examine individual patterns across themes
• Subcode emergent themes
1
Item Generation
• Structured items regarding gender ideology and relationship beliefs
Relationship Theme Extraction
• Use extracted items to create new themes for coding
Regression Design
• Use qualitative results to design model predicting relationship outcomes
3
2
Structured Interviews (n = 305)
• Sexual and relationship outcomes, social networks, contextual factors
Cultural Consensus
• Extract highest loading items regarding gender and relationship beliefs
Multilevel Dyadic Regression Analysis
• Run model with quantitative sample (n = 305 respondents/ n = 655 relationships)
Integrative Results Interpretation
• Interpret quantitative results in light of qualitative coding
• Construct integrative causal model, informed by both data sets

Source: Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli and Suzanne L. Wenzel. 2013. "Monogamy on the Street: A Mixed Methods Study of Homeless Men." Journal of Mixed Methods Research 7:3–28.

What emerged from the qualitative and quantitative elements of the analysis was an understanding of the homeless men as often upholding in theory the ideology of being responsible providers and partners, but experiencing near constant frustration at being unable to live up to these widely shared ideals.

> There's something on my forehead that says homeless or not good enough or

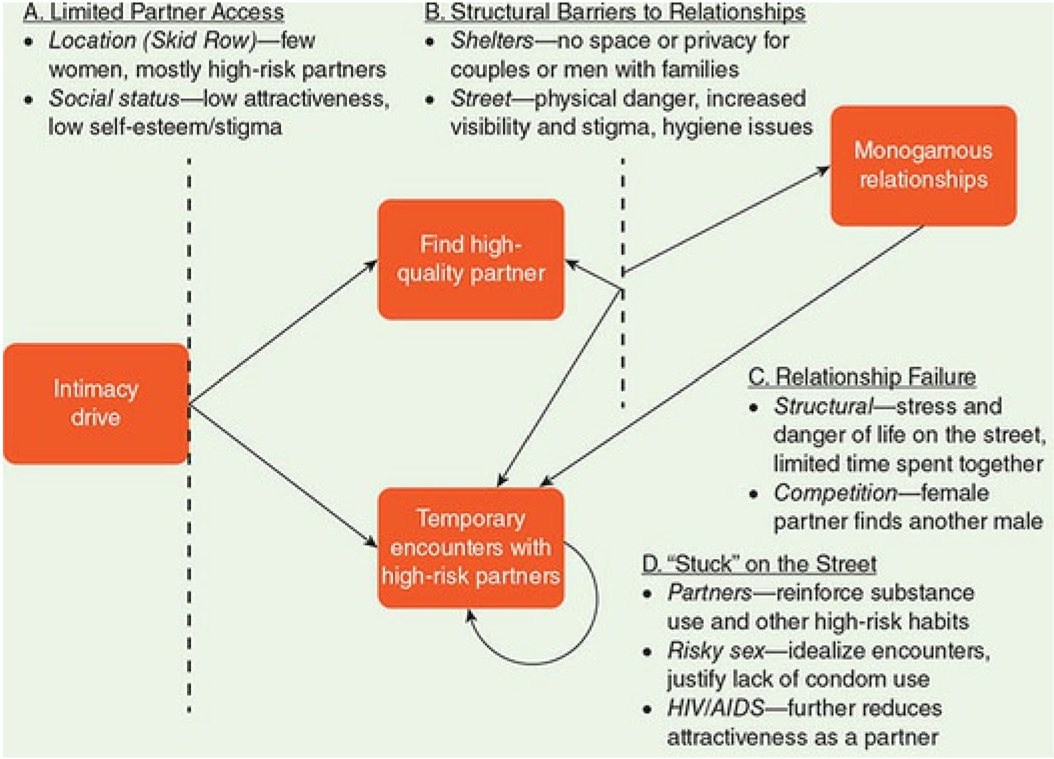
> income bracket, you know what I mean? . . . Not so much what you're wearing, but where you're living. . . . They'll cut you off real quick. (Brown et al. 2013:336).

The quantitative data indicated the men were less likely to be monogamous if their partners were also homeless or abusing substances, but more likely to be monogamous if their partners were involved in the same social network (Brown et al. 2013:338).

The diagram in Exhibit 12.4 displays Brown et al.'s complete explanatory model of relationships.

> Overall, our analyses illustrated how men maintained hopes and dreams of idealized, committed relationships despite (for the most part) being unable to realize these dreams while living on the street. They also showed how these idealized hopes and dreams in some cases lead [sic] to or were used to justify risk behaviors. In particular, analyses pointed to a mixture of social and structural barriers to men actualizing a normatively desired (but infrequently attained), low-risk behavioral pathway—being in committed, monogamous relationships. (Brown et al. 2013:341)

Exhibit 12.4 Holistic Model of High-Risk Sex Among Homeless Men


A. Limited Partner Access
• Location (Skid Row)—few women, mostly high-risk partners
• Social status—low attractiveness, low self-esteem/stigma
B. Structural Barriers to Relationships
• Shelters—no space or privacy for couples or men with families
• Street—physical danger, increased visibility and stigma, hygiene issues
Monogamous relationships
Find high-quality partner
Intimacy drive
C. Relationship Failure
• Structural—stress and danger of life on the street, limited time spent together
• Competition—female partner finds another male
Temporary encounters with high-risk partners
D. "Stuck" on the Street
• Partners—reinforce substance use and other high-risk habits
• Risky sex—idealize encounters, justify lack of condom use
• HIV/AIDS—further reduces attractiveness as a partner

Source: Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli and

Suzanne L. Wenzel. 2013. “Monogamy on the Street: A Mixed Methods Study of Homeless Men.” Journal of Mixed Methods Research 7:3–28.

Strengths and Limitations of Mixed Methods

Combining qualitative and quantitative methods within one research project can strengthen the project's design by enhancing measurement validity, generalizability, causal validity, or authenticity. At the same time, combining methods creates challenges that may be difficult to overcome and ultimately limit the extent to which these goals are enhanced. I will illustrate these strengths and limitations with examples from studies introduced in this chapter and from the Boston McKinney Project, the major mixed-methods study of housing alternatives for homeless persons diagnosed with severe mental illness that was funded by the National Institute of Mental Health and the Department of Housing and Urban Development (Schutt 2011b).

Measurement validity is enhanced when questions to be used in a structured quantitative survey are first refined through qualitative cognitive interviewing or focus groups. After quantitative measures are used, qualitative measures can be added to clarify the meaning of numerical scores. You learned about this advantage of combining methods in Chapter 8, on survey research. Alternatively, quantitative measures can provide a more reliable indicator of the extent of variation between respondents that has already been described based on naturalistic observations or qualitative interviews. You learned in Chapter 10 how qualitative observations or interview material can be categorized and coded so that repeated instances can be counted. You will learn in Chapter 14 how online sources can provide complementary quantitative and qualitative data that can be analyzed with mixed-methods designs (Salmons 2016:124).

In the News Research in the News: Why Women Don't Report Sexual Harassment

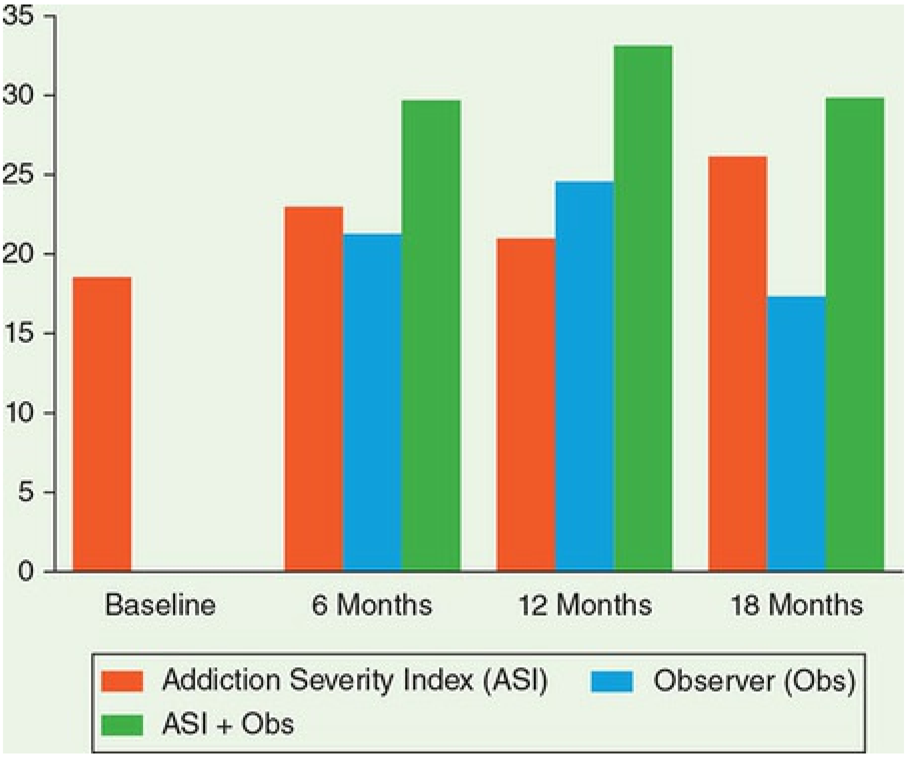
For Further Thought?

Sexual harassment in the workplace is all too often in the news, and those news articles are often followed with questions about why women didn’t report the alleged harassment when it first happened. How often are incidents of harassment not reported? To answer this question, Lilia Cortina, at the University of Michigan, and Jennifer Berdahl, at the University of British Columbia Sauder School of Business, analyzed data collected in 55 surveys about workplace incidents. Overall, about 25% of women report having experienced sexual harassment, while 50% have had specific experiences like inappropriate touching. However, only a quarter to a third report the incident at work, and just 2%–13% file a formal complaint. Reasons for nonreporting often focus on fears of retaliation.

1. Propose a qualitative method and a quantitative method for investigating sexual harassment in the workplace.
2. How could a mixed-methods design help determine whether there is a causal effect of company policies on the frequency of harassment?

News source: Miller, Claire Caine. 2017. “Why Women Often Don’t Report Sexual Harassment at Work.” *The New York Times,* April 11, p. B2.

Exhibit 12.5 Self-Report and Observer Ratings of Substance Abuse, by Follow-Up



Source: Adapted from Goldfinger, Stephen M., Russell K. Schutt, Larry J. Seidman, Winston M. Turner, Walter E. Penk, and George S. Tolomiczenko. "Self-Report and Observer Measures of Substance Abuse Among Homeless Mentally Ill Persons in the Cross-Section and Over Time." The Journal of Nervous and Mental Disease 184(11):667–672. Copyright © 1996, Wolters Kluwer, Lippincott, Williams & Wilkins. Reprinted with permission.

Measurement validity is also enhanced when measures using different methods result in a similar picture. If people are observed to do what they say they do, then the validity of a self-report measure is strengthened by measurement triangulation. The two mixed-methods studies of union democracy I introduced earlier in the chapter (Lipset et al. 1956; Schutt 1986) achieved this goal in their assessment of the bases of union factions when their quantitative survey data led to a description of union factional divisions that was similar to that resulting from observations of union meetings and qualitative interviews with factional leaders. McCarter's (2009) qualitative interviews on juvenile justice decisions corroborated the impression that emerged from quantitative data of racial bias in decisions.

But what if different measures do not reinforce each other? Did some respondents forget what they have done, or are they purposely not disclosing their activities? Are interview respondents trying to "look good" in interviews by inflating their rate of participation in socially desirable activity or by minimizing their rate of participation in socially undesirable behavior (Brenner 2012)?

My use of observational data about substance use in group homes in addition to quantitative survey data in *Homelessness, Housing, and Mental Illness* (Schutt 2011b) provides an interesting example of how to use mixed methods to understand the bases of a partial failure of triangulation. Our quantitative measures included both self-report measures of substance abuse and observer (case manager) ratings of substance abuse using a structured form. As indicated in Exhibit 12.5, these different quantitative measures did not entirely agree. Some substance abusers did not report substance use that the case managers observed, and the case managers did not observe all the substance abuse problems. We termed those who did not report substance abuse during a period that they were observed to abuse substances as "nondisclosers" (Goldfinger et al. 1996:671).

Observational notes recorded by project ethnographers helped explain the discrepant findings using these different measures. Some substance users would not acknowledge a problem that other project participants complained about:

> Tenants who were abusing alcohol and drugs often did not acknowledge any problem. One drinker explained that he "lost his toes [to frostbite] because he was on the street, not because he was drinking." Another argued that "the drinking cures mental illness" and another "says he is drinking because people aggravate him and don't give him the respect he deserves." . . . By contrast, tenants who did not have a substance abuse history or who were maintaining their sobriety often complained to staff that they felt threatened by others who were drinking. When one tenant insisted in a meeting that there was not a house drinking problem, another tenant retorted, "I'll disagree with that. Half the people in this house are drunk half the time." (Schutt 2011b:128)

In this example, qualitative data suggested that the failure of alternative quantitative measures to successfully triangulate resulted from unwillingness of some of those who abused substances to acknowledge a drinking problem. The result of using mixed methods was thus better understanding of the social reality underlying the individual behavior being measured. What at first might appear to be a limitation of mixed methods—the failure of alternative measures to converge, or triangulate—could instead be understood as an important insight because mixed methods were used.

The most common way that causal, or internal, validity is strengthened with a mixed-methods design is when qualitative interviews or observations are used to explore the mechanism involved in a causal effect. My mixed-methods analysis of the value of group and independent living options for people who had been homeless and diagnosed with severe mental illness demonstrates also how ethnographic data can help explain causal effects identified with an experimental design. Exhibit 12.6 displays the quantitative association between lifetime substance abuse—a diagnosis recorded on a numerical scale that was made on the basis of an interview with a clinician—and housing loss—another quantitative indicator from service records (Schutt 2011b:135).

The ethnographic notes recorded in the group homes revealed orientations and processes that helped explain the pronounced association between substance abuse and housing loss (Schutt 2011b):

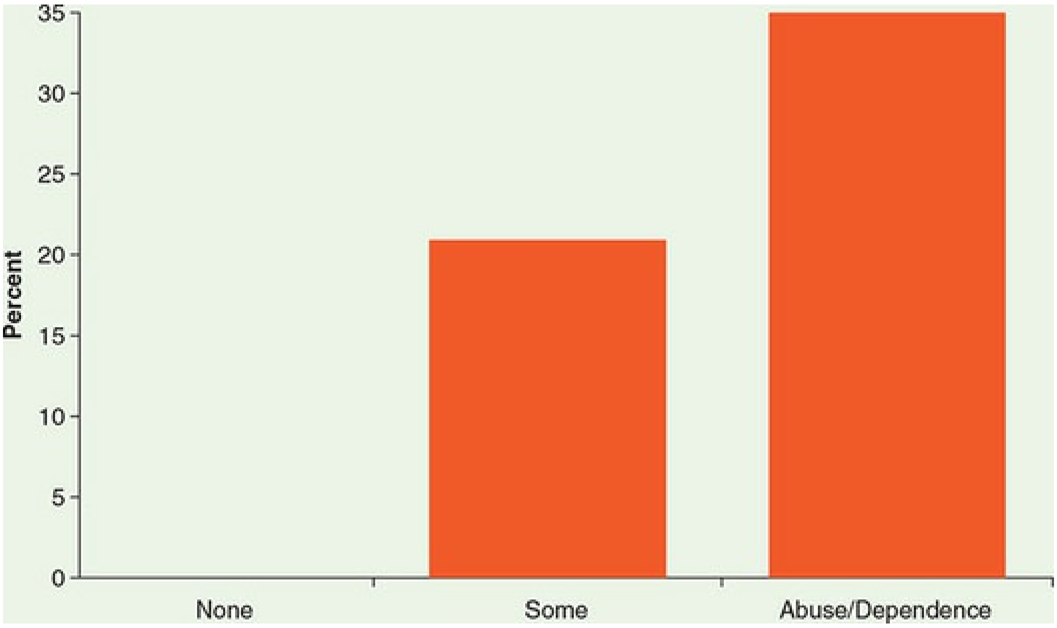
> The time has come where he has to decide once and for all to drink or not. . . . Tom has been feeling "pinned to the bed" in the morning. He has enjoyed getting high with Sammy and Ben, although the next day is always bad. . . . Since he came back from the hospital Lisandro has been acting like he is taunting them to throw him out by not complying with rules and continuing to drink. . . . (pp. 131, 133)

In this way, my analysis of the quantitative data reveals *what* happened, and my analysis of the ethnographic data helps to understand *why*. The same could be said for the way that qualitative interviews allowed Testa et al. (2011) to develop insights into how excessive alcohol consumption led to the quantitative association they identified between drinking and rape. It proved to be well worth the effort to collect qualitative data that would help interpret the results of the quantitative research.

A mixed-methods design can also improve external validity when a quantitative study is repeated in different contexts. Qualitative comparisons between these different contexts can then help make sense of the similarities and differences between outcomes across these contexts and thus help identify the conditions for the effects.

It was possible to make this type of qualitative comparison of quantitative study outcomes with the five experimental projects funded by the National Institute of Mental Health to identify the value of enhanced housing and services for reducing homelessness among homeless persons diagnosed with serious mental illness (our project in Boston was one of these five) (Schutt et al. 2009). Exhibit 12.7 displays the quantitative comparison for the impact of enhanced housing (the treatment) between these five projects. The results show that enhanced housing reduced the time spent homeless in four of the five projects. The one project that was an exception to this rule was in New York City and involved a focus on moving homeless persons off the streets into a shelter, which is an improvement in residential status but still counts as being homeless (and see Schutt 2011b:250–251). Interpretation of such multisite results must always be considered carefully, given the differences in project design that inevitably occur when the experiment is adapted to the specific conditions of each site.

Exhibit 12.6 Substance Abuse and Housing Loss in Group Homes

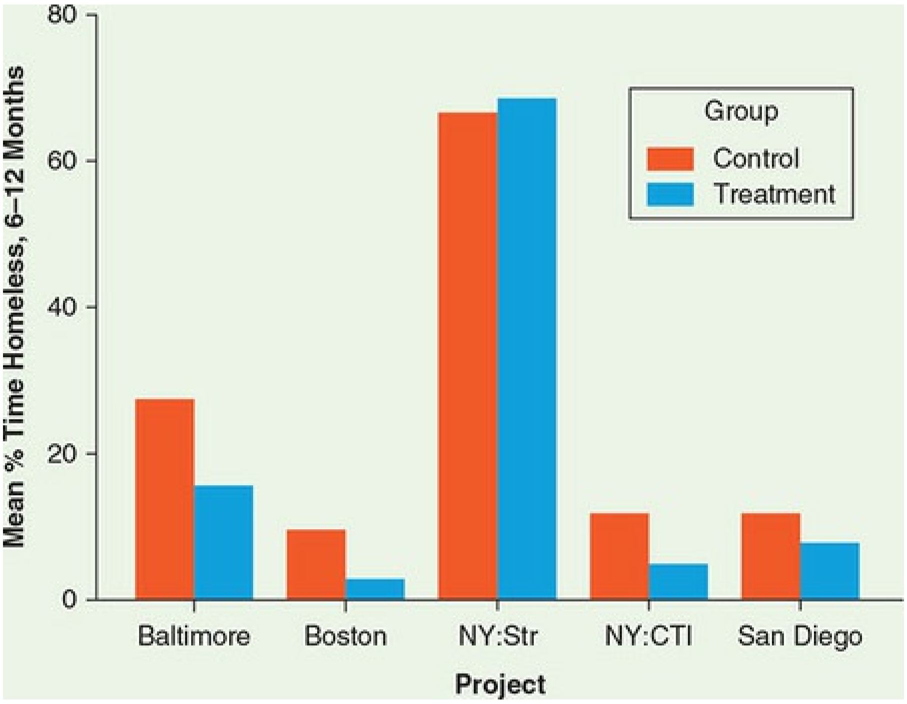


Source: Reprinted by permission of the publisher from Homelessness, Housing, and Mental Illness by Russell K. Schutt, with Stephen M. Goldfinger, p. 135. Cambridge, Mass.: Harvard University Press. Copyright © 2011 by the President and Fellows of Harvard College.

The generalizability of qualitative findings can also be improved when a representative sample developed with a quantitative design is used to identify cases to study more intensively with qualitative methods. This was the approach used by Zhong and Arnett (2014) when they selected Chinese women for qualitative interviews from among those who had participated in their structured survey. More commonly, the combination of methods to improve generalizability occurs in a research program in which the two studies are conducted sequentially, most often by different researchers.

Mixed methods also facilitate achieving the goal of authenticity within a research project that also seeks to achieve measurement validity, generalizability, or causal validity. Brown et al.'s (2013) complex mixed-method study of homeless men leads to a conclusion that seems much more authentic than would have been the case if they had used just one method. Their analysis moved them away from simplistic assumptions about the men's conceptions of appropriate gender roles to a deeper understanding of their gender role aspirations in the context of limited means for achieving them.

Exhibit 12.7 Mean Total Time Homeless by Project



It is naïve to think of mixed methods as simply overcoming problems that can occur when a research design is limited to one method. It is not always clear how best to compare qualitative and quantitative findings or how to interpret the discrepancies that arise after such comparison (Morgan 2014:66–81). These complexities, with unanticipated elements and insights, results in what some have called a certain “messiness” in mixed-methods research (Plano Clark and Ivankova 2016:276–277). We can’t be certain that differences in findings mean that deficits have been uncovered in one method—some substance abusers do not disclose their abuse in answer to questions—or that the two methods are really answering different research questions—for example, what is the association between drinking and rape on campus, compared with how often does drinking result in incapacitation prior to a rape?

Careers and Research

Amanda Aykanian, Research Associate, Advocates for Human Potential

Amanda Aykanian majored in psychology at Framingham State University and found that she enjoyed the routine and organization of research. She wrote an undergraduate thesis to answer the research question "How does the way in which course content is presented affect students' feelings about the content and the rate at which they retain it?"

After graduating, Aykanian didn't want to go to graduate school right away; instead she wanted to explore her interests and get a sense of what she could do with research. Advocates for Human Potential (AHP) was the last research assistant (RA) job for which Aykanian applied. Her initial tasks as an RA at AHP ranged from taking notes, writing agendas, and assembling project materials to entering research data, cleaning data, and proofing reports. As she contributed more to project reports, she began to think about data from a more theoretical standpoint.

During 7 years at AHP, Aykanian has helped lead program evaluation research, design surveys and write survey questions, conduct phone and qualitative interviews, and lead focus groups. Her program evaluation research almost always uses a mixed-methods approach, so Aykanian has learned a lot about how qualitative and quantitative methods can complement each other. She has received a lot of on-the-job training in data analysis and has learned how to think about and write a proposal in response to federal funding opportunities.

Aykanian was promoted to research associate and describes her current role as part program evaluation coordinator and part data analyst. She has also returned to graduate school, earning a master's degree in applied sociology and then starting a PhD program in social welfare.

Mixed methods also create extra challenges for researchers because different types of expertise are required for effective use of quantitative and qualitative methods. Recruiting multiple researchers for a project who then work as a team from conception to execution of the project may be the best way to overcome this limitation, but it may be difficult to mesh researchers with different backgrounds and perspectives (Plano Clark and Ivankova 2016:225, 228). The researchers also have to acknowledge in planning the study timetable that the time required for collection, coding, and analysis of qualitative data can challenge a quantitative researcher's expectation of more rapid progress.

Despite these challenges, the different types of evidence produced in a mixed-methods investigation can strengthen overall confidence in the research findings and result in a more holistic understanding of the social world. Weaver-Hightower (2013) captured these advantages in his methodological reflections on his mixed-method study of influences on public policy in Australia:

> Overall, this mixed-methods process, moving iteratively from qualitative to quantitative back to qualitative, established well the political and ideological influences on the policy makers and their "policy." The quantitative and qualitative methods were highly integrated—that is, the whole of the findings exceeded the sum of the individual quantitative and qualitative parts . . . because the quantitative procedures in some cases solidified, and in other cases challenged, my qualitative impressions of influence that were hard-won by hanging around, talking to people, and reading about the subject. Indeed, I was surprised by several influentials identified using the embedded quantitative phase. Several groups and individuals, like Canberra Grammar School or Boys in Focus, largely escaped my notice before the mixed-methods approach. Likewise, without the qualitative case descriptions and the negative and positive case analyses, I would have been less able to understand why certain groups and actors were influential. And, regarding negative cases, the qualitative methods were still necessary to identify some influentials not netted by the mixed methods. In the end, then, the methods were integrated because both were necessary to find and evaluate influentials. (p. 17)

Ethics and Mixed Methods

Researchers who combine methods must be aware of the ethical concerns involved in using each of the separate methods, but there are also some ethical challenges that are heightened in mixed-methods projects. One special challenge is defining the researcher's role in relation to the research participants. Every researcher creates an understanding about his or her role with research participants (Mertens 2012). Researchers using quantitative methods often define themselves as outside experts who design a research project and collect research data using objective procedures that are best carried out without participant involvement. By contrast, qualitative researchers often define themselves as engaging in research in some type of collaboration with the community or group they are studying, with much input from their research participants into the research design and the collection and analysis of research data.

A researcher using mixed methods cannot simply adopt one of these roles: A researcher needs some degree of autonomy when designing quantitative research plans, but a researcher will not be able to collect intensive qualitative data if participants do not accord her or him some degree of trust as an insider. The challenge is compounded by the potential for different reactions of potential participants to the different roles. Authorities who control access to program clients or employees or to community members may be willing to agree to a structured survey but not to a long-term engagement with researchers as participant observers, so that a mixed-methods project that spans programs, communities, or other settings may involve a biased sampling for the qualitative component. Natalia Luxardo, Graciela Colombo, and Gabriela Iglesias (2011) confronted this challenge in their study of Brazilian family violence services and as a result focused their qualitative research on one service that supported the value of giving voice to their service recipients.

Weighing both roles and the best combination of them is critical at the outset of a mixed-methods project, although the dilemma will be lessened if a project uses different researchers to lead the quantitative and qualitative components. In our study of housing alternatives, for example, a team of ethnographers collected data on activities in the group homes while research assistants supervised by a different leader collected the project's quantitative data (Schutt 2011b).

Complex mixed-methods projects in which quantitative surveying is interspersed with observational research or intensive interviews may also require renegotiation of participant consent to the particular research procedures at each stage. As stated by Chih Hoong Sin (2005),

> Different stages and different components of research may require the

> negotiation of different types of consent, some of which may be more explicit than others. Sampling, contact, re-contact, and fieldwork can be underpinned by different conceptualization and operationalization of “informed consent.” This behooves researchers to move away from the position of treating consent-seeking as an exercise that only occurs at certain points in the research process or only for certain types of research. Consent-seeking should not be thought of merely as an event. (p. 290)

In the qualitative component of their study of Brazilian victims of domestic violence, Luxardo and her colleagues (2011) adopted a flexible qualitative interviewing approach to allow participants to avoid topics they did not want to discuss:

> We tried to consider what was important for that adolescent during the interview and, many times, we had to reframe the content of the encounters according to the expectations they had. So, if they were not willing to share during an interview but still had complaints, doubts, or comments to share, we tried to focus on those instead of subtly directing the talk to the arena of the research interests. Moreover, we noticed that some adolescents (most of them migrants from Bolivia) did not feel at ease sharing that kind of information about their lives with a stranger, so we tried not to invade their intimacy by being culturally sensitive; if they did not want to talk, they did not have to do so. (p. 996)

Conclusions

A research project that is designed to answer multiple research questions and investigate a complex social setting often requires a mixed-methods design. Of course, to some extent the complexity of the social world always exceeds what can be captured successfully with one method, but the challenges increase as our questions become more numerous and our social settings more complex. You have learned in this chapter about the different perspectives on combining methods, the different ways of mixing methods, and the challenges that arise when mixing methods. No matter what your methodological preference is at this point, increased understanding of these issues in mixed methods will improve your own research practice and your ability to critique the research of others.

I conclude with my justification for the mixed methods used in the project that was the foundation for my recent book about a very ambitious set of research questions in a complex and changing social setting (Schutt 2011b:284):

> *Homelessness, Housing, and Mental Illness* describes who participated in the Boston McKinney Project, what they wanted, and what they did; it evaluates whether living in group homes or independent apartments influenced participants' desires and actions; it explores how participants interacted with each other and whether they differed in their responses to the same stimuli. For these reasons and more, our project required systematic research methods—not just one method, such as conducting a survey, but multiple methods: different methods to answer different questions and to provide alternative perspectives on similar questions. . . . Even the best and most appropriate research methods do not entirely solve the problem of perspective. When we use social science methods, we see farther and probe deeper than we do in our everyday lives; however, no method—whether used alone or in combination with others—gives us perfect vision or infallible insight. Every method we use for investigating the social world will overlook some processes, distort some realities, and confuse some issues.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Complex mixed-methods design 473
Embedded mixed-methods design 471
Integrated mixed-methods design 470
Mixed methods 467
Multitrait–multimethod matrix 467
Paradigm wars 468
Research program with mixed methods 469
Staged mixed-methods design 472
Triangulation 467

Highlights

- Researchers use mixed methods because a single method may not represent adequately the complexity of the social world that they are trying to understand.
- Sociologists and other social scientists used mixed methods in community studies as long ago as the 1930s, but their popularity declined with the rapid growth of quantitative methods in the postwar period and during the subsequent period of the paradigm wars. Mixed methods have become much more popular in the past three decades.
- Campbell and Fiske (1959) proposed the use of a multitrait–multimethod matrix to improve testing for measurement validity.
- The paradigm war between those who favor qualitative and quantitative methods emerged from inflexibly linking positivist philosophy to quantitative methods and constructivist philosophy to qualitative methods.
- Some researchers who use mixed methods sidestep consideration of paradigms in favor of selecting methods on the basis of what seems most appropriate for a given research question.
- Mixed methods combine qualitative and quantitative methods in a systematic way in an investigation of the same or related research questions.
- Types of mixed methods can be distinguished by the priority given to one method or the other, with the basic distinction being between designs that give equal priority to the two methods and those that prioritize one method over the other. Types of methods can also be distinguished by sequencing, with sequential designs using one method before the other and concurrent designs using both methods at the same time.
- In a research program, qualitative and quantitative methods are used sequentially and given equal priority.
- In an integrated mixed-methods design, qualitative and quantitative methods are used concurrently and both are given equal priority.
- In an embedded mixed-methods design, qualitative and quantitative methods are used concurrently but one has priority.
- In a staged mixed-methods design, one method is used in a sequence before the other, but they have equal priority.
- In a complex mixed-methods design, at least one of the four primary designs is combined with another of these designs, or one is used multiple times.
- Measures are said to be triangulated when different types of measures of the same phenomenon give consistent results.
- Mixed methods heighten some ethical concerns, including developing access procedures appropriate to both qualitative and quantitative roles, and renegotiating consent for different components of the research.

Discussion Questions

1. Testa et al. (2011:237) claim, "Mixed-methods research can benefit any area of study," but also argue that this approach is particularly appropriate for the study of violence against women. What makes mixed methods more appropriate for research on the problem of violence against women, in their opinion? What about in your own opinion? Are there other areas that you think are more suited to the use of mixed methods? Explain your reasoning.
2. Testa describes her own training as quantitative and then highlights some experiences that led her to integrate qualitative methods into her research (Testa et al. 2011). Would you describe your own training in research methods so far as primarily quantitative or qualitative? Have you had any experiences that led you to consider the "other" methodology?
3. Which of the four types of mixed methods do you feel is likely to be useful for investigating the social world? Would you favor more single-method or more mixed-methods studies? Explain your reasoning.
4. Has this textbook led you to favor either qualitative or quantitative methods, or has it encouraged you to use mixed methods when possible? As a textbook, why has it had this effect?
5. Consider how ready you feel to design a mixed-methods research project. Do you think mixed-methods researchers should generally try to collaborate with another researcher who specializes in the methodology they are not so familiar with? Or should researchers seek to become experts in multiple methods so that they can combine them in research projects that they direct themselves?

Practice Exercises

1. Select a public setting in which there are many people on a regular basis, such as a sports venue, a theater lobby, a coffee shop, or a popular public park. Observe as an ethnographer for 30 minutes on one day and then write up your notes. Before your next visit, a day later, develop a systematic observation schedule on which to record observations in the same setting in a structured manner. Record observations using the structured observation form on another day, but in the same place and at about the same time of day. Compare the data you collected with your ethnographic notes and with your systematic observation notes.
2. Read one of the four SAGE mixed-methods research articles (available on the study site) highlighted in this chapter. What would have been lost if this study had been a single-method study?
3. Complete the interactive exercises about mixed methods on the book's study site.

Ethics Questions

1. Would you prefer your role as a researcher to be that of an objective outside expert, or as a collaborator helping local participants solve problems they are concerned about? What conditions would make the role of outside expert preferable? Explain.
2. Should separate consent forms and processes be used for the qualitative and quantitative components of a mixed-methods project? What would be the advantages and drawbacks of this approach?

Web Exercises

1. Read the abstracts for 5–10 recent articles in SAGE's *Journal of Mixed Methods Research* (http://journals.sagepub.com/home/mmr). Do you find any general similarities in the research questions addressed? Does it seem that any social research could benefit from the use of mixed methods or that some research questions are more appropriate for investigation with mixed methods than others? Explain your reasoning.
2. Read the National Institute of Health's online report on *Best Practices of Mixed Methods Research in the Health Sciences* (https://obssr.od.nih.gov/training/mixed-methods-research/). What additional desirable features of mixed-methods research are identified?

Video Interview Questions

Listen to the researcher interview with Dana Hunt for Chapter 12 at edge.sagepub.com/schutt9e.

1. Why was this specific research study challenging?
2. How did the researchers come up with the "counterfactual" component of the study?

SPSS Exercises

Could qualitative data help you understand quantitative associations between variables?

1. Generate the cross-tabulations of CONARMY, CONBUS, CONCLERG, CONEDUC, CONFED, CONFINAN, CONJUDGE, CONLABOR, CONLEGIS, CONMEDIC by EDUCR in the GSS2016 or GSS2016x data set (with column percents).
2. What do you learn about the relation between education and confidence in social institutions from these crosstabs?
3. What questions would you like to include in a qualitative interview with some GSS respondents in order to improve your understanding of the education–confidence association? Explain.

Developing a Research Proposal

Add a mixed-methods design for your proposed study. Pick the type of mixed method that seems most likely to help answer the research question (see Exhibit 3.10, #13 to #17).

1. Explain why it will be advantageous to give priority to either qualitative or quantitative methods, or to give them equal priority.
2. Explain why you think it advantageous to sequence the methods, or not.

Chapter 13 Evaluation and Policy Research

Research That Matters, Questions That Count
History of Evaluation Research
Careers and Research
Evaluation Basics
Research in the News: No-Cost Talk Therapy?
Questions for Evaluation Research
  Needs Assessment
  Evaluability Assessment
  Process Evaluation
  Impact Analysis
  Efficiency Analysis
Design Decisions
  Black Box Evaluation or Program Theory
  Researcher or Stakeholder Orientation
  Quantitative or Qualitative Methods
  Simple or Complex Outcomes
  Groups or Individuals
Policy Research
Ethics in Evaluation
Conclusions

Drug Abuse Resistance Education (D.A.R.E.) is offered in elementary schools, middle schools, and high schools across the United States and in 49 other countries (D.A.R.E. 2014). Since its inception in 1983 as an innovative effort of the Los Angeles Police Department, D.A.R.E. has grown to include 15,000 police officers and as many as 75% of U.S. school districts (Berman and Fox 2009). For parents worried about drug abuse among youth and for many concerned citizens, the program has immediate appeal. It brings a special police officer into the schools once a week to talk to students about the hazards of drug abuse and to establish a direct link between local law enforcement and young people. Although there are many substance abuse prevention programs, none have achieved the popularity or developed the infrastructure of D.A.R.E.

Research That Matters, Questions That Count

Evaluation research on the Drug Abuse Resistance Education (D.A.R.E.) program in schools has long raised questions about its impact on drug abuse. However, there have been positive findings about the impact of program participation on students' attitudes toward the police. Amie Schuck at the University of Illinois at Chicago focused on this aspect of program impact in her evaluation of D.A.R.E. She was able to use for this evaluation data already collected in a large randomized experiment that had tested the impact of D.A.R.E. in 12 pairs of urban and suburban schools in Illinois. Students' attitudes toward police had been measured with their answers to five questions asked in seven waves of data collection over a 7-year period.

Schuck found that student attitudes toward the police became considerably more negative from the 5th and 6th grades, when the study began, to the 11th and 12th grades, when the study concluded, although by this point the decline was reduced. These changes in attitudes are similar to what has been found in other studies of youth attitudes toward the police. However, participation in the D.A.R.E. program delayed the decline in attitudes toward the police and then was associated with improved attitudes toward the police. This association was particularly strong for African American youth. Schuck highlights several implications of her study for criminal justice policy.

1. Schuck reviews several theories that try to explain the decline in youth attitudes toward the police. What theory do you suppose would be most persuasive? Explain your reasoning.
2. How would you design a "process evaluation" of the effect of D.A.R.E. on youth attitudes toward the police? What specific methods would you propose to investigate the way in which the program had this effect?

In this chapter, you will learn about different types of evaluation research and the decisions that researchers must make when they plan to evaluate the impact of social programs. You will also learn more about the history of research designed to evaluate D.A.R.E. By the end of the chapter, you will understand why agencies so often require evaluation research to test the effectiveness of government programs, and you will have a much firmer basis for answering the questions I have posed. After you finish the chapter, test yourself by reading the 2013 *Journal of Research in Crime and Delinquency* article by Amie Schuck at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 13 at edge.sagepub.com/schutt9e.

Schuck, Amie M. 2013. "A Life-Course Perspective on Adolescents' Attitudes to Police: DARE, Delinquency, and Residential Segregation." *Journal of Research in Crime and Delinquency* 50(4):579–607.

But does it work? Do students who participate in D.A.R.E. education become more resistant to the use of illicit drugs? Are they less likely to use illicit drugs while they are enrolled in the program or, more important, in the years after they have finished the program? Do students benefit in other ways from participation in D.A.R.E.? Are there beneficial effects for schools and communities? Although the idea of providing students with information about the harmful effects of substance abuse has intuitive appeal, the history of evaluation research about D.A.R.E. drives home an important point: To know whether social programs work, and to understand how they work, we have to evaluate them systematically and fairly, whether we personally think the program sounds like a good idea or not.

This point was not lost on those charged with funding or administering D.A.R.E. The U.S. Department of Justice, which had funded the initial expansion of D.A.R.E. as a national program, also paid for a program evaluation by Susan Ennett, Christopher Ringwalt, and Robert Flewelling at the Research Triangle Institute in North Carolina, and Nancy Tobler at the State University of New York in Albany. Ennett and her colleagues (1994) located eight quantitative studies of the effects of D.A.R.E. that had used an experimental or quasi-experimental design with a focus on a particular state or locality. Across these studies, D.A.R.E. had no effect on drug or alcohol use at the time students completed D.A.R.E., although it led to a small reduction in tobacco use (and see Ringwalt et al. 1994; West and O'Neal 2004). D.A.R.E. participants did improve their knowledge about substance use, as well as their social skills related to resisting substance abuse, attitudes toward the police, attitudes about drug use, and self-esteem, but these positive attitudinal effects were less than those identified in evaluations of other types of substance abuse prevention programs. A 6-year randomized field experiment of D.A.R.E.'s effectiveness in a sample of Illinois schools also found no long-term beneficial effects (Rosenbaum and Hanson 1998). As a result, some school districts stopped using D.A.R.E. (Rosenbaum 2007).

Yet this was not the end of the story. Federal officials convened a meeting of researchers and D.A.R.E. administrators to consider program changes and more research. As a result, the Robert Wood Johnson Foundation funded substance abuse researcher Zili Sloboda at the University of Akron to develop a new educational approach and a rigorous evaluation of its long-term impact (Berman and Fox 2009). Surprisingly, Sloboda and her colleagues (2009) found that this new program, Take Charge of Your Life, actually led to increased use of alcohol and cigarettes and no change in marijuana use. So D.A.R.E. administrators rejected that approach, adopted a different model ("keepin' it REAL"), and retooled once again (Toppo 2002; West and O'Neal 2004).

> Gone is the old-style approach to prevention in which an officer stands behind a podium and lectures students in straight rows. New D.A.R.E. officers are trained as "coaches" to support kids who are using research-based refusal strategies in high-stakes peer-pressure environments. (D.A.R.E. 2008)

Of course, the "new D.A.R.E." is now being evaluated, too. Sorry to say, one early quasi-experimental evaluation in 17 urban schools, funded by D.A.R.E. America, found no effect of the program on students' substance use (Vincus et al. 2010). Some researchers have concluded that the program should simply be ended, while others have concluded that some communities have had good reasons for continuing to offer the program (Berman and Fox 2009; Birkeland, Murphy-Graham, and Weiss 2005).

This may seem like a depressing way to begin a chapter on evaluation research. If, like me, you have a child who enjoyed D.A.R.E., or were yourself a D.A.R.E. student, you may know how appealing the program can be and how important its message is. But that should help to drive home the key point: Government agencies and other bodies that fund social programs must invest the necessary resources to evaluate their effectiveness, no matter how appealing they seem on their surface. Resources are too scarce to spend millions of dollars on a national program that does not achieve its intended goals. As we review more of the story of evaluation research on D.A.R.E. and other programs in this chapter, you will also learn that program evaluations can give us much more insight into the complexity of the social world. Evaluation findings can convince us that our preconceptions about human motivations or program processes need to be revised; they can identify differences in how different people react to the same program and how the same program may have different effects in different social contexts; and they can alert us to the importance of some program outcomes—both harmful and beneficial—that we may have overlooked. Moreover, the efforts of federal officials to encourage reexamination of and change in the D.A.R.E. program demonstrate that program evaluations can have real impacts on public policy.

In this chapter, you will read about a variety of social program evaluations as I introduce the evaluation research process, illustrate the different types of evaluation research, highlight alternative approaches, and review ethical concerns. You will learn in this chapter that current major debates about social policies such as health care, homelessness, domestic violence, and drug abuse are each being informed by important major evaluation projects and that broad "policy research" projects can help define new directions for public action. You should finish the chapter with a much better understanding of how the methods of applied social research can help improve society.

History of Evaluation Research

Evaluation research is not a method of data collection, like survey research or experiments, nor is it a unique component of research designs, like sampling or measurement. Instead, evaluation research is social research that is conducted for a distinct purpose: to investigate social programs (e.g., substance abuse treatment programs, welfare programs, criminal justice programs, or employment and training programs). For each project, an evaluation researcher must select a research design and a method of data collection that are useful for answering the particular research questions posed and appropriate for the particular program investigated. Policy research, introduced later in the chapter, leads to recommendations based on the results of multiple projects.

The development of evaluation research as a major enterprise followed on the heels of the expansion of the federal government during the Great Depression and World War II. Large Depression-era government outlays for social programs stimulated interest in monitoring program output, and the military effort in World War II led to some of the necessary review and contracting procedures for sponsoring evaluation research. In the 1960s, criminal justice researchers began to use experiments to test the value of different policies (Orr 1999:24). New government social programs of the 1960s often came with evaluation requirements attached, and more than 100 contract research and development firms began operation in the United States between 1965 and 1975 (Dentler 2002; Rossi and Freeman 1989:34). The first evaluation research journal—*Evaluation Review*—was launched by SAGE in 1976 (Fox, Grimm, and Caldeira 2016:16).

The New Jersey Income Maintenance Experiment was the first large-scale randomized experiment to test social policy in action. Designed in 1967, the New Jersey experiment randomly assigned 1,300 families to different income support levels to test the impact of cash transfers to the working poor on their work effort. It was soon followed by even larger experiments to test other income maintenance questions, most notably the Seattle–Denver Income Maintenance Experiment (Orr 1999:24–26). The number of social experiments like this continued to increase in subsequent years (Greenberg, Shroder, and Onstott 1999:159). By 2001, there were more than 10,000 citations to what appeared to be randomized trials in social, psychological, educational, and criminological research (Petrosino et al. 2001). By 2017, the Campbell Collaboration research archive (www.campbellcollaboration.org), a project begun by evaluation researchers in 1999, contained 304 systematic reviews of social research evaluation results for sets of investigations ranging in focus from crime and justice and social welfare to education and international development.

Government requirements and popular concern with the benefit derived from taxpayer-funded programs continue to stimulate evaluation research. The Community Mental

Health Act Amendments of 1975 (Public Law 94–63) required quality assurance reviews, which often involved evaluation-like activities (Patton 2002:147–151), and the Government Performance and Results Act of 1993 required some type of evaluation of all government programs (U.S. Office of Management and Budget 2002). In 1998, the federal Safe and Drug-Free Schools and Communities Act was reauthorized with the proviso that all programs to be funded had to be research based (Petrosino et al. 2001). At century's end, the federal government was spending about $200 million annually on evaluating $400 billion in domestic programs, and the 30 major federal agencies had between them 200 distinct evaluation units (Boruch 1997). In 1999, the new Governmental Accounting Standards Board urged that more attention be given to "service efforts and accomplishments" in standard government fiscal reports (Campbell 2002). This same period saw increasing requirements for performance indicators, performance audits, and accreditation reviews of public programs, which can themselves stimulate interest in formal evaluation research designed to improve performance (Fox et al. 2016:5–6). Nonprofit organizations like Britain's Social Care Institute for Excellence (www.scie.org.uk) encourage adoption of evidence-based practices.

Evaluation Basics

Exhibit 13.1 illustrates the process of evaluation research as a simple systems model. First, clients, customers, students, or some other persons or units—cases—enter the program as inputs. (You'll notice that this model treats programs like machines, with people functioning as raw materials to be processed.) Students may begin a new school program, welfare recipients may enroll in a new job training program, or crime victims may be sent to a victim advocate. The resources and staff a program requires are also program inputs.

Next, some service or treatment is provided to the cases. This may be attendance in a class, assistance with a health problem, residence in new housing, or receipt of special cash benefits. The program process may be simple or complicated, short or long, but it is designed to have some impact on the cases.

The direct product of the program's service delivery process is its output. Program outputs may include clients served, case managers trained, food parcels delivered, or arrests made. The program outputs may be desirable in themselves, but they primarily indicate that the program is operating.

Program outcomes indicate the impact of the program on the cases that have been processed. Outcomes can range from improved test scores or higher rates of job retention to fewer criminal offenses and lower rates of poverty. Any social program is likely to have multiple outcomes—some intended and some unintended, some positive and others that are viewed as negative.

Variation in both outputs and outcomes, in turn, influences the inputs to the program through a feedback process. If not enough clients are being served, recruitment of new clients may increase. If too many negative side effects result from a trial medication, the trials may be limited or terminated. If a program does not appear to lead to improved outcomes, clients may go elsewhere.

Evaluation research is simply a systematic approach to feedback: It strengthens the feedback loop through credible analyses of program operations and outcomes. Evaluation research also broadens this loop to include connections to parties outside the program itself. A funding agency or political authority may mandate the research, outside experts may be brought in to conduct the research, and the evaluation research findings may be released to the public, or at least the funders, in a formal report.

The evaluation process can be understood only in relation to the interests and perspectives of program stakeholders. Stakeholders are those individuals and groups who have some basis of concern with the program. They might be clients, staff, managers, funders, or the public. The board of a program or agency, the parents or spouses of clients, the foundations that award program grants, the auditors who monitor program spending, the members of Congress—each is a potential stakeholder, and each has an interest in the outcome of any program evaluation. Some may fund the evaluation; some may provide research data; and some may review, or even approve, the research report (Martin and Kettner 1996:3). Who the program stakeholders are and what role they play in the program evaluation will have tremendous consequences for the research.

Inputs: The resources, raw materials, clients, and staff that go into a program.

Program process: The complete treatment or service delivered by the program.

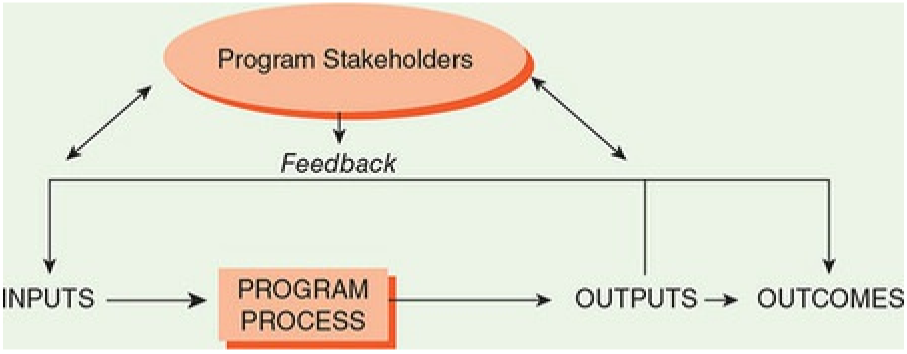
Outputs: The services delivered or new products produced by the program process.

Outcomes: The impact of the program process on the cases processed.

Feedback: Information about service delivery system outputs, outcomes, or operations that can guide program input.

Stakeholders: Individuals and groups who have some basis of concern with the program.

Exhibit 13.1 A Model of Evaluation



Source: Adapted from Martin and Kettner (1996).

Can you see the difference between evaluation research and traditional social science research (Posavac and Carey 1997)? Unlike explanatory social science research, evaluation research is not designed to test the implications of a social theory; instead, it is designed to answer a basic question: What is the program's impact? Process evaluation often uses qualitative methods like traditional social science does, but unlike exploratory research, the goal is not to induce a broad theoretical explanation for what is discovered. Instead, the question is, how does the program do what it does? Unlike social science research, the researchers cannot design evaluation studies simply in accord with the highest scientific standards and the most important research questions; instead, the program stakeholders set the agenda. But there is no sharp boundary between the two. In their attempt to explain how and why a program has an impact, and whether the program is needed, evaluation researchers often bring social theories into their projects but for immediately practical aims.

Questions for Evaluation Research

Evaluation projects can focus on several questions related to the operation of social programs and the impact they have:

- Is the program needed?
- Can the program be evaluated?
- How does the program operate?
- What is the program's impact?
- How efficient is the program?

The specific methods used in an evaluation research project depend partly on which of these foci the project has.

Needs Assessment

Is a new program needed or an old one still required? Is there a need at all? A needs assessment attempts to answer these questions with systematic, credible evidence. Need may be assessed by social indicators, such as the poverty rate or the level of home ownership; by interviews of local experts, such as school board members or team captains; by surveys of populations in need; or by focus groups composed of community residents (Rossi and Freeman 1989).

Needs assessment is not as easy as it sounds (Posavac and Carey 1997). Whose definitions or perceptions should be used to shape our description of the level of need? How will we deal with ignorance of need? How can we understand the level of need without understanding the social context from which that level of need emerges? (Short answer to that one: We can’t!) What, after all, does *need* mean in the abstract? We won’t really understand what the level of need is until we develop plans for implementing a program in response to the identified needs.

Needs assessment: A type of evaluation research that attempts to determine the needs of some population that might be met with a social program.

In the News Research in the News: No-Cost Talk Therapy?

For Further Thought?

England is conducting a national experiment to improve the treatment of depression, anxiety, and other common mental illnesses with talk therapy. The national government is currently spending about $500 million in an effort to ensure that residents who have one of these illnesses are diagnosed and referred for appropriate—and free—treatment, in a way that minimizes costs. Those who call their local Healthy Minds program are interviewed on the phone for an hour and then referred for more therapy sessions on the phone, or in group or individual therapy sessions, depending on their level of need. Progress is tracked with standard questionnaires filled out each week and tracked anonymously. Treatment may continue for several weeks, months, or longer, but early evaluations indicate demand is strong, rates of recovery are good, and so many millions are being saved in lost time at work due to illness.

1. What hypothesis would you propose to test about the value of providing no-cost talk therapy and what research design would you suggest using to test it?
2. Describe a possible research project about no-cost talk therapy using the policy research approach described in this chapter.

News source: Carey, Benedict. 2017. "England's Mental Health Experiment: No-Cost Talk Therapy." *The New York Times,* July 24.

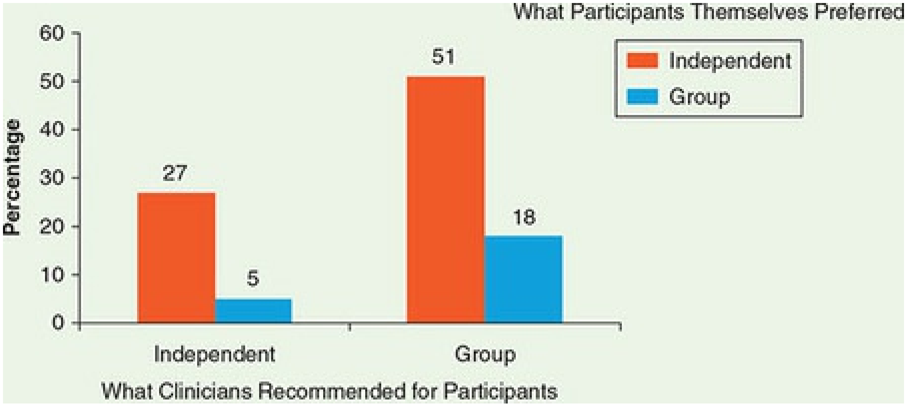
The results of the Boston McKinney Project reveal the importance of taking a multidimensional approach to the investigation of need. The Boston McKinney Project evaluated the merits of providing formerly homeless mentally ill persons with staffed group housing rather than with individual housing (Schutt 2011b). In a sense, you can think of the whole experiment as involving an attempt to answer the question "What type of housing do these persons 'need'?" Our research team first examined this question at the start of the project by asking each project participant which type of housing he or she wanted (Schutt and Goldfinger 1996) and by independently asking two clinicians to estimate which of the two housing alternatives would be best for each participant (Goldfinger and Schutt 1996).

Exhibit 13.2 displays the findings. The clinicians recommended staffed group housing for 69% of the participants, whereas most of the participants (78%) sought individual housing.

There was no correspondence between the housing recommendations of the clinicians and the housing preferences of the participants (who did not know what the clinicians had recommended for them). So which perspective reveals the level of need for staffed group housing versus individual housing?

Yet another perspective on housing needs is introduced by the project's outcomes. Individuals assigned to the group housing were somewhat more successful in retaining their housing than were those who were assigned to individual housing, and this differential success rate grew in the years after the project's end (Schutt 2011b:247). Does this therefore reveal that these homeless mentally ill persons "needed" group housing more than they needed individual housing, despite their preference? What should we make of the fact that the participants who preferred individual housing were more likely to lose their housing during the project, whereas the participants whom the clinicians had rated as ready for independent living were less likely to lose their housing (Schutt 2011b:248)? And what should we make of the fact that whether or not participants received the type of housing the clinicians recommended or that they themselves preferred made no difference to the likelihood of their losing their housing during the project (Schutt 2011b:248)? Does this mean that neither initial preferences nor clinician recommendations tell us about the need for one or the other type of housing, only about the risk of losing whatever housing they were assigned to?

Exhibit 13.2 Type of Residence: Preferred and Recommended



Source: Based on Goldfinger and Schutt (1996).

The methodological lesson here is that in needs assessment, as in other forms of evaluation research, it is a good idea to use multiple indicators. You can also see that there is no absolute definition of *need* in this situation, nor is there likely to be in any but the most simplistic evaluation projects. A good evaluation researcher will do his or her best to capture different perspectives on need and then help others make sense of the results.

A wonderful little tale, popular with evaluation researchers, reveals the importance of thinking creatively about what people need:

> The manager of a 20-story office building had received many complaints about the slowness of the elevators. He hired an engineering consultant to propose a solution. The consultant measured traffic flow and elevator features and proposed replacing the old with new ones, which could shave 20 seconds off the average waiting time. The only problem: It cost $100,000. A second consultant proposed adding 2 additional elevators, for a total wait time reduction of 35 seconds and a cost of $150,000. Neither alternative was affordable. A third consultant was brought in. He looked around for a few days and announced that the problem was not really the waiting times, but boredom. For a cost of less than $1,000, the manager had large mirrors installed next to the elevators so people could primp and observe themselves while waiting for an elevator. The result: no more complaints. Problem solved. (Witkin and Altschuld 1995:38)

Evaluability Assessment

Evaluation research will be pointless if the program itself cannot be evaluated. Yes, some type of study is always possible, but a study specifically to identify the effects of a particular program may not be possible within the available time and resources. So researchers may conduct an evaluability assessment to learn this in advance, rather than expend time and effort on a fruitless project.

Why might a social program not be evaluable? Several factors may be involved:

- Management only wants to have its superior performance confirmed and does not really care whether the program is having its intended effects. This is a very common problem.
- Staff are so alienated from the agency that they don't trust any attempt sponsored by management to check on their performance.
- Program personnel are just "helping people" or "putting in time" without any clear sense of what the program is trying to achieve.
- The program is not clearly distinct from other services delivered from the agency and so can't be evaluated by itself (Patton 2002:164).

Because evaluability assessments are preliminary studies to "check things out," they often rely on qualitative methods. Program managers and key staff may be interviewed in depth, or program sponsors may be asked about the importance they attach to different goals. The evaluators may then suggest changes in program operations to improve evaluability. They may also use the evaluability assessment to "sell" the evaluation to participants and sensitize them to the importance of clarifying their goals and objectives. If the program is judged to be evaluable, knowledge gleaned through the evaluability assessment is used to shape evaluation plans.

Complex community initiatives can be particularly difficult to evaluate due to the evolving nature of the intervention as it is adapted to community conditions, an often broad range of outcomes, and—many times—the absence of a control or comparison community (Fox et al. 2016:43). The President's Family Justice Center (FJC) Initiative was such an initiative begun by President George W. Bush to plan and implement comprehensive domestic violence services that would provide "one stop shopping" for victims in need of services. In 2004, the National Institute of Justice contracted with Abt Associates in Cambridge, Massachusetts, to assess the evaluability of 15 pilot service programs that had been awarded a total of $20 million and to develop an evaluation plan. In June 2005, Abt researchers Meg Townsend, Dana Hunt, Caity Baxter, and Peter Finn reported on their evaluability assessment.

Abt's assessment began with conversations to collect background information and perceptions of program goals and objectives from those who had designed the program. These conversations were followed by a review of the grant applications submitted by each of the 15 sites and phone conversations with site representatives. Site-specific data collection focused on the project's history at the site, its stage of implementation, staffing plans and target population, program activities and stability, goals identified by the site's director, apparent contradictions between goals and activities, and the state of data systems that could be used in the evaluation. Exhibit 13.3 shows the resulting logic model that illustrates the intended activities, outcomes, and impacts for the Alameda County, California, program. Although they had been able to begin the evaluability assessment process, Townsend and her colleagues concluded that in the summer of 2005, none of the 15 sites were far enough along with their programs to complete the assessment.

Evaluability assessment: A type of evaluation research conducted to determine whether it is feasible to evaluate a program's effects within the available time and resources.

Process Evaluation

What actually happens in a social program? The New Jersey Income Maintenance Experiment was designed to test the effect of some welfare recipients receiving higher payments than others (Kershaw and Fair 1976). Did that occur? In the Minneapolis experiment on the police response to domestic violence (Sherman and Berk 1984), police officers were to either arrest or warn individuals accused of assaulting their spouses on the basis of a random selection protocol, unless they concluded that they must override the experimental assignment to minimize the risk of repeat harm. Did the police officers follow this protocol? How often did they override it because of concerns about risk? Questions like these about program implementation must be answered before it is possible to determine whether the program's key elements had the desired effect. Answers to such program implementation questions are obtained through process evaluation—research to investigate the process of service delivery.

Process evaluation is even more important when more complex programs are evaluated. Many social programs comprise multiple elements and are delivered over an extended period, often by different providers in different areas. Because of this complexity, it is quite possible that the program as delivered is not the same for all program recipients or consistent with the formal program design.

The evaluation of D.A.R.E. by Research Triangle Institute researchers Ringwalt and colleagues (1994:7) included a process evaluation with three objectives:

1. To assess the organizational structure and operation of representative D.A.R.E. programs nationwide
2. To review and assess the factors that contribute to the effective implementation of D.A.R.E. programs nationwide
3. To assess how D.A.R.E. and other school-based drug prevention programs are tailored to meet the needs of specific populations

The process evaluation (they called it an *implementation assessment*) was an ambitious research project in itself, with site visits, informal interviews, discussions, and surveys of D.A.R.E. program coordinators and advisers. These data indicated that D.A.R.E. was operating as designed and was running relatively smoothly. As shown in Exhibit 13.4, drug prevention coordinators in D.A.R.E. school districts rated the program components as more satisfactory than did coordinators in school districts with other types of alcohol and drug prevention programs.

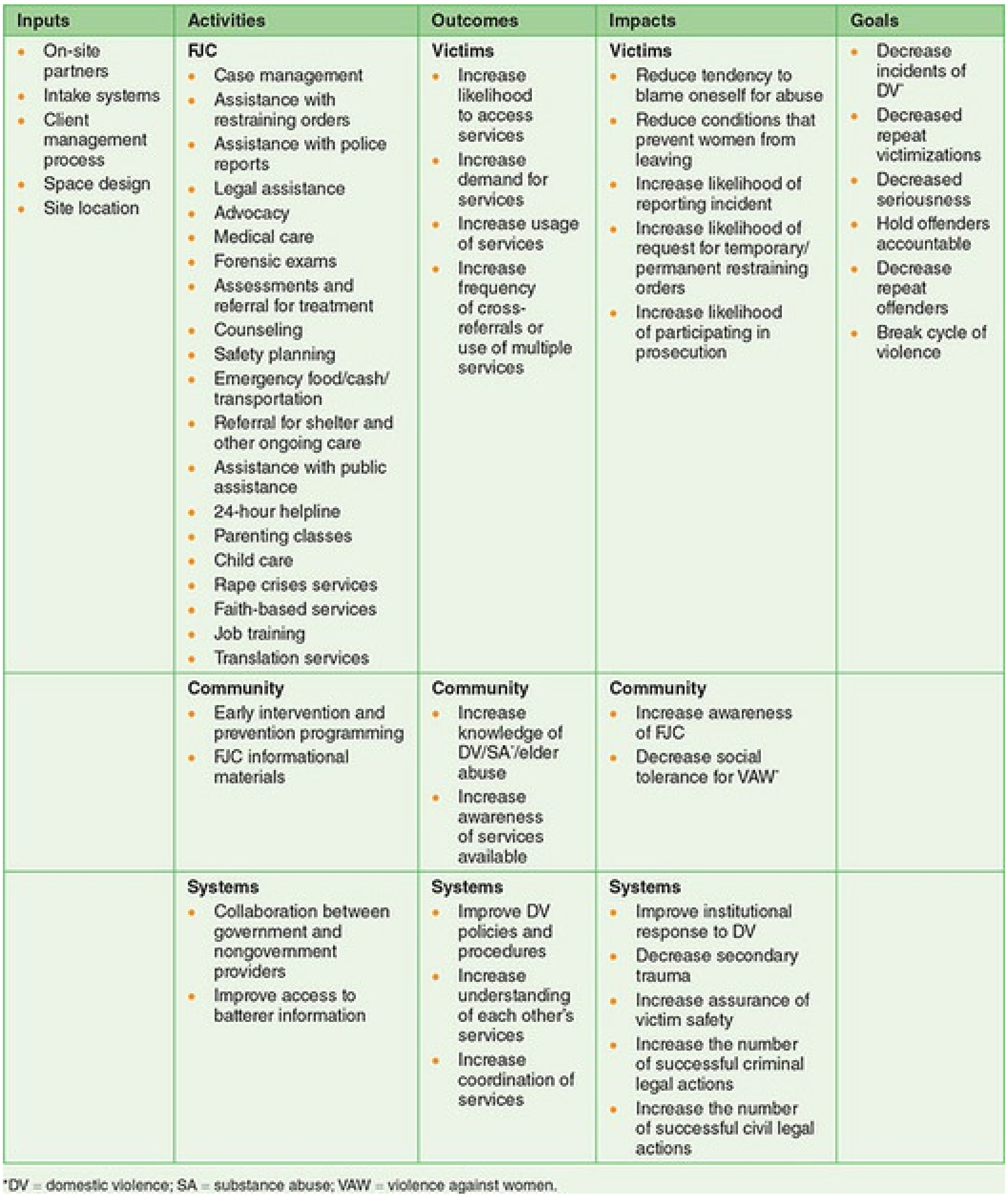
Process evaluation also can be used to identify the specific aspects of the service delivery process that have an impact. This, in turn, will help explain why the program has an effect and which conditions are required for these effects. (In Chapter 6, I described this as identifying the causal mechanism.) For instance, implementation problems identified in D.A.R.E. site visits included insufficient numbers of officers to carry out the program as planned and a lack of Spanish-language D.A.R.E. books in a largely Hispanic school. Classroom observations indicated engaging presentations and active student participation (Ringwalt et al. 1994:58). Sloboda et al.'s (2009) evaluation of the trial *Take Care of Your Life* curriculum found that D.A.R.E. officers delivered all of the lessons and 73% of the planned content in those lessons—a rate of program delivery that is better than that in many large-scale efforts (Sloboda et al. 2009:8).

Process analysis of this sort can show how apparent findings may be incorrect. The apparently disappointing results of the Transitional Aid Research Project (TARP) provide an instructive lesson of this sort. TARP was a social experiment designed to determine whether financial aid during the transition from prison to the community would help released prisoners find employment and avoid returning to crime. Two thousand participants in Georgia and Texas were randomized to receive either a particular level of benefits over a particular period or no benefits at all (the control group). Initially, it seemed that the payments had no effect: The rate of subsequent arrests for both property and nonproperty crimes was not affected by the TARP treatment condition.

Process evaluation: Evaluation research that investigates the process of service delivery.

But this wasn't all there was to it. Peter Rossi tested a more elaborate causal model of TARP effects, summarized in Exhibit 13.5 (Chen 1990). Participants who received TARP payments had more income to begin with and so had more to lose if they were arrested; therefore, they were less likely to commit crimes. However, TARP payments also created a disincentive to work and therefore increased the time available in which to commit crimes. Thus, the positive direct effect of TARP (more to lose) was cancelled out by its negative indirect effect (more free time).

Exhibit 13.3 Alameda Family Justice Center Logic Model



Inputs	Activities	Outcomes	Impacts	Goals
• On-site partners • Intake systems • Client management process • Space design • Site location	**FJC** • Case management • Assistance with restraining orders • Assistance with police reports • Legal assistance • Advocacy • Medical care • Forensic exams • Assessments and referral for treatment • Counseling • Safety planning • Emergency food/cash/transportation • Referral for shelter and other ongoing care • Assistance with public assistance • 24-hour helpline • Parenting classes • Child care • Rape crises services • Faith-based services • Job training • Translation services	**Victims** • Increase likelihood to access services • Increase demand for services • Increase usage of services • Increase frequency of cross-referrals or use of multiple services	**Victims** • Reduce tendency to blame oneself for abuse • Reduce conditions that prevent women from leaving • Increase likelihood of reporting incident • Increase likelihood of request for temporary/permanent restraining orders • Increase likelihood of participating in prosecution	• Decrease incidents of DV* • Decreased repeat victimizations • Decreased seriousness • Hold offenders accountable • Decrease repeat offenders • Break cycle of violence
	Community • Early intervention and prevention programming • FJC informational materials	**Community** • Increase knowledge of DV/SA*/elder abuse • Increase awareness of services available	**Community** • Increase awareness of FJC • Decrease social tolerance for VAW*	
	Systems • Collaboration between government and nongovernment providers • Improve access to batterer information	**Systems** • Improve DV policies and procedures • Increase understanding of each other's services • Increase coordination of services	**Systems** • Improve institutional response to DV • Decrease secondary trauma • Increase assurance of victim safety • Increase the number of successful criminal legal actions • Increase the number of successful civil legal actions	

*DV = domestic violence; SA = substance abuse; VAW = violence against women.

Source: Meg Townsend, Dana Hunt, Caity Baxter, and Peter Finn. 2005. Interim Report: Evaluability Assessment of the President's Family Justice Center Initiative. Cambridge, MA: Abt Associates Inc. Reprinted with permission.

The term formative evaluation may be used instead of process evaluation when the evaluation findings are used to help shape and refine the program (Rossi and Freeman 1989). Formative evaluation procedures that are incorporated into the initial development of the service program can specify the treatment process and lead to changes in recruitment procedures, program delivery, or measurement tools (Patton 2002:220). They can provide a strong foundation for managers as they develop the program (Fox et al. 2016:10)

Formative evaluation: Process evaluation that is used to shape and refine program operations.

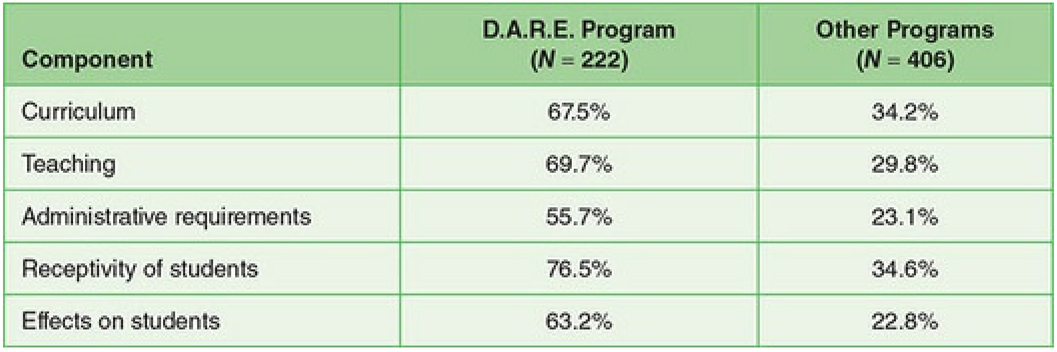
You can see the formative element in the following excerpt from the report on Ohio's Assisted Living Medicaid Waiver Program by a research team from the Scripps Gerontology Center and Miami University (Applebaum et al. 2007). The program is designed to allow disabled elderly persons to move into assisted living facilities that maximize autonomy rather than more expensive nursing homes.

> In order for the state to develop a viable assisted living program, a plan to increase provider participation is critical. . . . The perspectives of residents, and their families, providers, case managers, and representatives of the Department of Aging and Health will be needed to refine the program as it develops. . . . Every program needs to evolve and therefore a solid structure to identify and implement necessary changes will be crucial for long-term program success. (p. 31)

Process evaluation can employ a wide range of indicators. Program coverage can be monitored through program records, participant surveys, community surveys, or users versus dropouts and ineligibles. Service delivery can be monitored through service records completed by program staff, a management information system maintained by program administrators, or reports by program recipients (Rossi and Freeman 1989).

Qualitative methods are often a key component of process evaluation studies because they can be used to elucidate and understand internal program dynamics—even those that were not anticipated (Patton 2002:159; Posavac and Carey 1997). Qualitative researchers may develop detailed descriptions of how program participants engage with each other, how the program experience varies for different people, and how the program changes and evolves over time. The goal is to develop a "bottom-up" view of the process, rather than the "top-down" view that emerges from official program records (Fox et al. 2016:64–66).

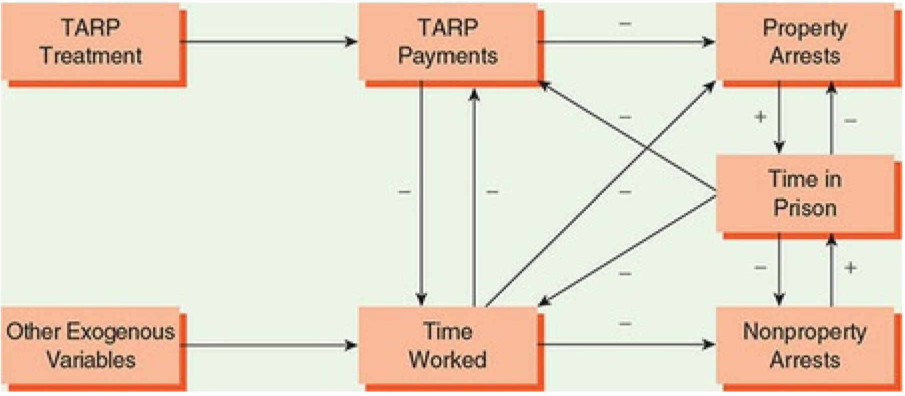
Exhibit 13.4 Components of D.A.R.E. and Other Alcohol and Drug Prevention Programs Rated as Very Satisfactory (%)



Component	D.A.R.E. Program (*N* = 222)	Other Programs (*N* = 406)
Curriculum	67.5%	34.2%
Teaching	69.7%	29.8%
Administrative requirements	55.7%	23.1%
Receptivity of students	76.5%	34.6%
Effects on students	63.2%	22.8%

Source: Ringwalt et al. (1994:58).

Exhibit 13.5 Model of TARP Effects


TARP Treatment
TARP Payments
Property Arrests
Time in Prison
Other Exogenous Variables
Time Worked
Nonproperty Arrests
–
–
–
–
–
+
–
–
–
–
+
–

Source: Chen (1990:210). Reprinted with permission from SAGE Publications, Inc.

Impact Analysis

The core questions of evaluation research are “Did the program work?” and “Did it have the intended result?” This part of the research is variously called impact analysis, impact evaluation, or summative evaluation. Formally speaking, impact analysis compares what happened after a program with what would have happened had there been no program.

Think of the program—a new strategy for combating domestic violence, an income supplement, whatever—as manipulating an independent variable and the result it seeks as a change in dependent variable. The D.A.R.E. program (independent variable), for instance, tries to reduce drug use (dependent variable). When the program is present, we expect less drug use. In a more elaborate study, we might have multiple values of the independent variable; for instance, we might look at *no program, D.A.R.E. program,* and *other drug/alcohol education* conditions and compare the results of each. In a complex intervention, it is important to measure program components carefully in order to determine whether specific components were responsible for any impacts identified (Fox et al. 2016: 84).

Elizabeth J. D’Amico and Kim Fromme’s (2002) study of a new Risk Skills Training Program (RSTP) is a good example of a more elaborate study. They compared the impact of RSTP on children aged 14–19 years with that of an abbreviated version of D.A.R.E. and with results for a control group. The impacts they examined included positive and negative alcohol expectancies (the anticipated effects of drinking) as well as perception of peer risk taking and actual alcohol consumption. They found that negative alcohol expectancies increased for the RSTP group in the posttest but not for the D.A.R.E. group or the control group, and weekly drinking and positive expectancies for drinking outcomes actually *increased* for the D.A.R.E. group or the control group by the 6-month follow-up but not for the RSTP group (pp. 568–570; see Exhibit 13.6).

Impact evaluation (or analysis): Analysis of the extent to which a treatment or other service has an effect; also known as summative evaluation.

As in other areas of research, an experimental design is the preferred method for maximizing internal validity—that is, for making sure your causal claims about program impact are justified. Cases are assigned randomly to one or more experimental treatment groups and to a control group so that there is no systematic difference between the groups at the outset (see Chapter 7). The goal is to achieve a fair, unbiased test of the program itself, so that differences between the types of people who are in the different groups do not influence judgment about the program’s impact. It can be a difficult goal to achieve because the usual practice in social programs is to let people decide for themselves whether they want to enter a program or not and to establish eligibility criteria that ensure that people who enter the program are different from those who do not (Boruch 1997). In either case, a selection bias is introduced.

Impact analyses that do not use an experimental design can still provide useful information and may be all that is affordable, conceptually feasible, or ethically permissible in many circumstances. Evaluation of the State Children's Health Insurance Program (SCHIP) provides an example. The U.S. Congress enacted SCHIP in 1997 to expand health insurance coverage for low-income uninsured children. The federal Centers for Medicare & Medicaid Services (CMS) then contracted with Mathematica Policy Research to evaluate SCHIP (Rosenbach et al. 2007:1).

Given the nature of SCHIP, children could not be assigned randomly to participate. Instead, the Mathematica researchers tracked the growth of enrollment in SCHIP and changes in the percentage of children without health insurance. You can see in Exhibit 13.7 that the percentage of uninsured children declined from 25.2 to 20.1 from 1997 to 2003, during which time enrollment in SCHIP grew from 0.7 million to 6.2 million children (Rosenbach et al. 2007:ES.2). SCHIP, Margo Rosenbach and her colleagues (2007) concluded,

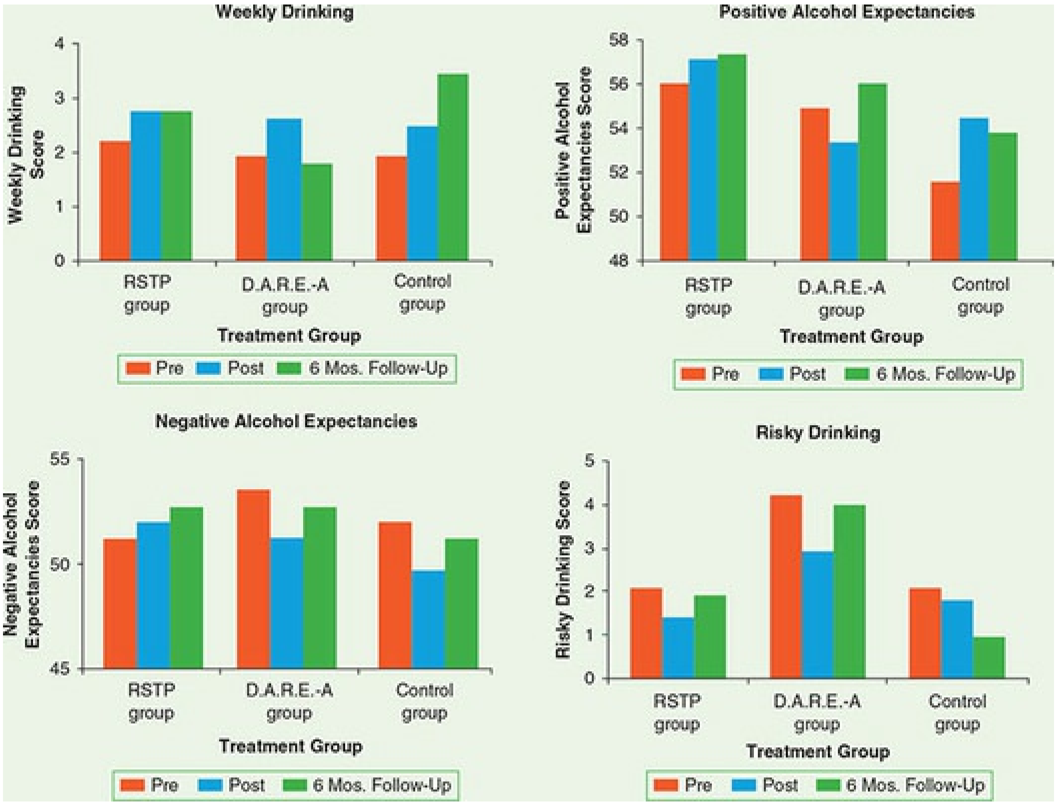
> provided a safety net for children whose families lost employer-sponsored coverage during the economic downturn [of 2000–2003] . . . [even as] nonelderly adults experienced a significant 2 percentage point increase in their overall uninsured rate. . . . In the absence of SCHIP . . . the number of uninsured children would have grown by 2.7 million, rather than declining by 0.4 million. (p. ES.7)

Of course, program impact may also be evaluated with quasi-experimental designs (see Chapter 7) or survey or field research methods. But if current participants who are already in a program are compared with nonparticipants, it is unlikely that the treatment group will be comparable with the control group. Participants will probably be a selected group, different at the outset from nonparticipants. As a result, causal conclusions about program impact will be on much shakier ground. For instance, when a study at New York's maximum-security prison for women concluded "Income Education [i.e., classes] Is Found to Lower Risk of New Arrest," the findings were immediately suspect: The research design did not ensure that the women who enrolled in the prison classes were similar to those who had not enrolled in the classes, "leaving open the possibility that the results were due, at least in part, to self-selection, with the women most motivated to avoid re-incarceration being the ones who took the college classes" (Lewin 2001a:A18).

Rigorous evaluations often lead to the conclusion that a program does not have the desired effect (Patton 2002:154). A program depends on political support for achieving its goals, and such evaluations may result in efforts to redesign the program (as with D.A.R.E.) or reduction or termination of program funding. The latter outcome occurred with the largest U.S. federal training program for the disadvantaged, the Job Training Partnership Act (JTPA). In 1995, partly because of negative results of evaluation research, Congress both restructured JTPA and cut its funding by more than 80% (Heckman, Hohmann, and Smith 2000:651).

But the JTPA experience also raises an important caution about impact analysis. The National JTPA Study used a rigorous experimental design to evaluate the JTPA program and concluded that it did not have the desired impact. However, a subsequent, more intensive analysis by the economist James Heckman and his colleagues (2000) indicated that the initial evaluation had overlooked two critical issues: (1) between 27% and 40% of control group members found some other training program to participate in while the research was being conducted and (2) between 49% and 59% of treatment group members dropped out of the program and so did not actually receive the classroom training (p. 660). After Heckman et al. accounted for these problems of control group "substitution bias" and treatment group dropout, they concluded that classroom training in JTPA had "a large positive effect on monthly earnings" for those completing training (p. 688). There is more of a basis for concern when program participants know that a special intervention is being tested and not everyone is receiving it. Staff delivering a program in the control condition could adopt some elements of the intervention being tested to try to help their clients, while program clients could become demoralized because they are not receiving the intervention (Fox et al. 2016:93). So it is important to design research carefully and consider all possible influences on program impact before concluding that a program should be terminated because of poor results.

Exhibit 13.6 Impact of RSTP, D.A.R.E.-A

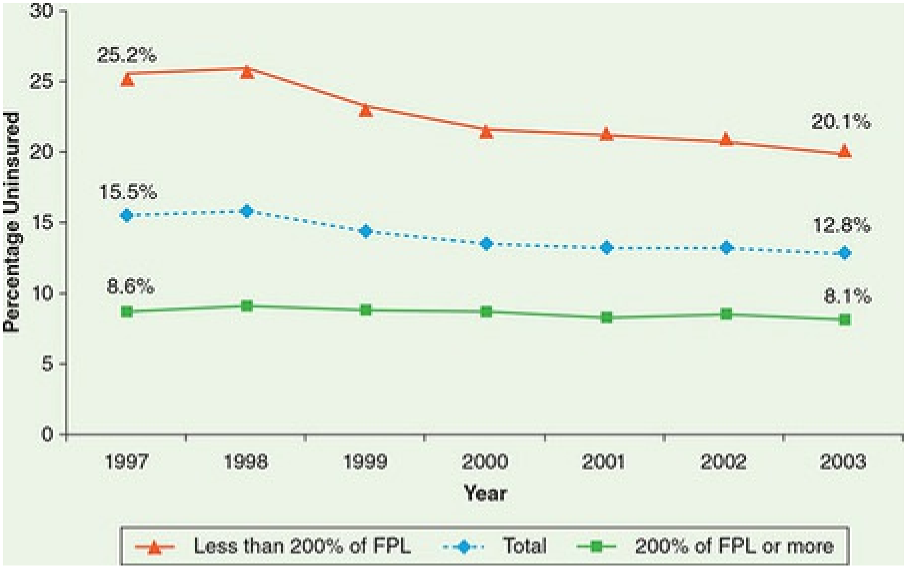


Source: Based on D’Amico and Fromme (2002:569).

Efficiency Analysis

Whatever the program's benefits, are they sufficient to offset the program's costs? Are the taxpayers getting their money's worth? What resources are required by the program? These efficiency questions can be the primary reason why funders require evaluation of the programs they fund. As a result, efficiency analysis, which compares program effects with costs, is often a necessary component of an evaluation research project.

Exhibit 13.7 Percentage of Children Under Age 19 Without Health Insurance by Poverty Level



Source: Margo Rosenbach, Carol Irvin, Angela Merrill, Shanna Shulman, John Czajka, Christopher Trenholm, Susan Williams, So Sasigant Limpa-Amara, and Anna Katz. 2007. National Evaluation of the State Children's Health Insurance Program: A Decade of Expanding Coverage and Improving Access: Final Report. Cambridge, MA: Mathematica Policy Research, Inc.

A cost–benefit analysis must identify the specific program costs and the procedures for estimating the economic value of specific program benefits. This type of analysis also requires that the analyst identify whose perspective will be used to determine what can be considered a benefit rather than a cost. Program clients will have a different perspective on these issues than will taxpayers or program staff. Exhibit 13.8 lists the factors that can be considered as costs or benefits in an employment and training program, from the standpoint of program participants, the rest of society, and society as a whole (the combination of program participants and the rest of society) (Orr 1999:224). Note that some anticipated impacts of the program, on welfare benefits and wage subsidies, are considered a cost to one group and a benefit to another group, whereas some are not relevant to one of the groups.

A cost–effectiveness analysis focuses attention directly on the program's outcomes rather than on the economic value of those outcomes. In a cost–effectiveness analysis, the specific costs of the program are compared with the program's outcomes, such as the number of jobs obtained, the extent of improvement in reading scores, or the degree of decline in crimes committed. For example, one result might be an estimate of how much it cost the program for each job obtained by a program participant.

Social science training often doesn't devote much attention to cost–benefit analysis, so it can be helpful to review possible costs and benefits with an economist or business school professor or student. Once the potential costs and benefits have been identified, they must be measured. This is a need highlighted in new government programs (Campbell 2002):

> The Governmental Accounting Standards Board's (GASB) mission is to establish and improve standards of accounting and financial reporting for state and local governments in the United States. In June 1999, the GASB issued a major revision to current reporting requirements ("Statement 34"). The new reporting will provide information that citizens and other users can utilize to gain an understanding of the financial position and cost of programs for a government and a descriptive management's discussion and analysis to assist in understanding a government's financial results. (p. 1)

In addition to measuring services and their associated costs, a cost–benefit analysis must be able to make some type of estimation of how clients benefited from the program. Normally, this will involve a comparison of some indicators of client status before and after clients received program services or between clients who received program services and a comparable group that did not.

A recent study of therapeutic communities (TCs) provides a clear illustration. A TC is a method for treating substance abuse in which abusers participate in an intensive, structured living experience with other addicts who are attempting to stay sober. Because the treatment involves residential support as well as other types of services, it can be quite costly. Are those costs worth it?

Stanley Sacks and colleagues (2002) conducted a cost–benefit analysis of a modified TC. In the study, 342 homeless, mentally ill chemical abusers were randomly assigned to either a TC or a "treatment-as-usual" comparison group. Employment status, criminal activity, and utilization of health care services were each measured for the 3 months before entering treatment and the 3 months after treatment. Earnings from employment in each period were adjusted for costs incurred by criminal activity and utilization of health care services.

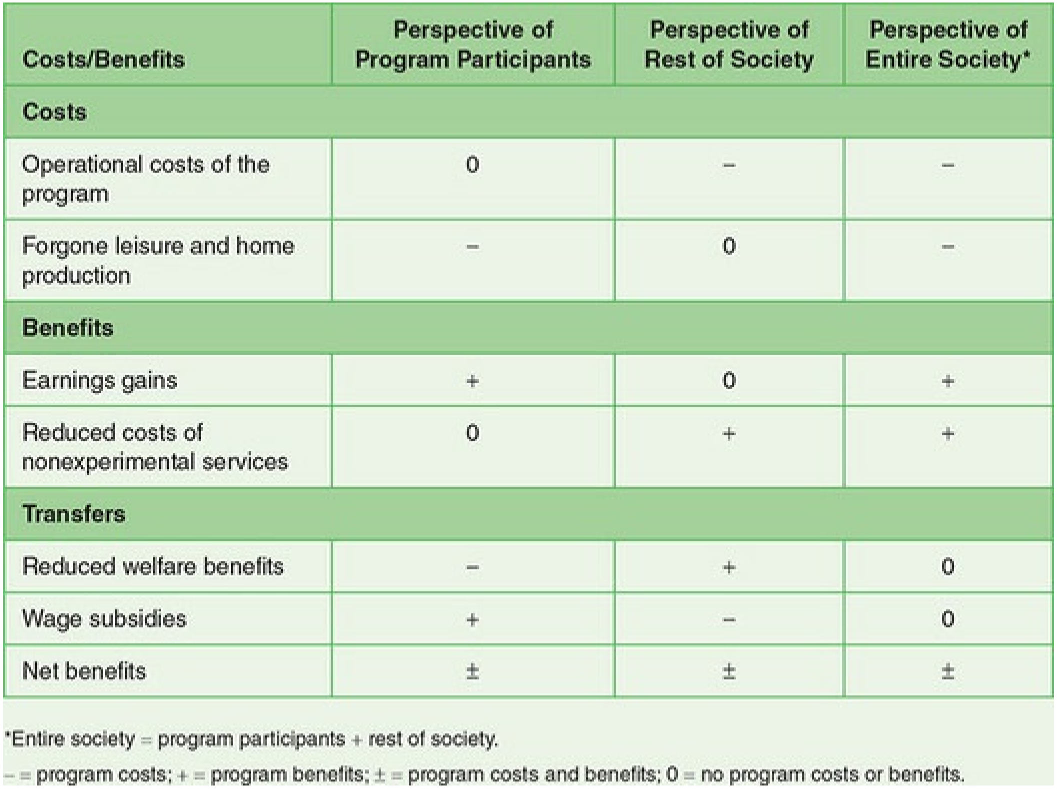
Efficiency analysis: A type of evaluation research that compares program costs with program effects. It can be either a cost–benefit analysis or a cost–effectiveness analysis.

Cost–benefit analysis: A type of evaluation research that compares program costs with the economic value of program benefits.

Cost–effectiveness analysis: A type of evaluation research that compares program costs with actual program outcomes.

Was it worth it? The average cost of TC treatment for a client was $20,361. In comparison, the economic benefit (based on earnings) to the average TC client was $305,273, which declined to $273,698 after comparing postprogram with preprogram earnings, and it was still $253,337 even after adjustment for costs. The resulting benefit–cost ratio was 13:1, although this ratio declined to only 5.2:1 after further adjustments (for cases with extreme values). Nonetheless, the TC program studied seems to have had a substantial benefit relative to its costs.

Exhibit 13.8 Conceptual Framework for Cost–Benefit Analysis of an Employment and Training Program



Costs/Benefits	Perspective of Program Participants	Perspective of Rest of Society	Perspective of Entire Society*
Costs			
Operational costs of the program	0	–	–
Forgone leisure and home production	–	0	–
Benefits			
Earnings gains	+	0	+
Reduced costs of nonexperimental services	0	+	+
Transfers			
Reduced welfare benefits	–	+	0
Wage subsidies	+	–	0
Net benefits	±	±	±

*Entire society = program participants + rest of society.

– = program costs; + = program benefits; ± = program costs and benefits; 0 = no program costs or benefits.

Source: Orr (1992:224, Table 6.5).

Design Decisions

Once we have decided on, or identified, the goal or focus of a program evaluation, there are still important decisions to be made about how to design the specific evaluation project. The most important decisions are the following:

- *Black box evaluation or program theory:* Do we care how the program gets results?
- *Researcher or stakeholder orientation:* Whose goals matter the most?
- *Quantitative or qualitative methods:* Which methods provide the best answers?
- *Simple or complex outcomes:* How complicated should the findings be?
- *Groups or individuals:* Assign groups or individuals to different programs?

Black Box Evaluation or Program Theory

The "meat and potatoes" of most evaluation research involves determining whether a program has the intended effect. If the effect occurred, the program has "worked"; if the effect didn't occur, then, some would say, the program should be abandoned or redesigned. In black box evaluation, the process by which a program has an effect on outcomes is often treated as a black box—that is, the focus of the evaluation researcher is on whether cases seem to have changed as a result of their exposure to the program, between the time they entered the program as inputs and when they exited the program as outputs (Chen 1990). The assumption is that program evaluation requires only the test of a simple input–output model, like that shown in Exhibit 13.1. There may be no attempt to open the black box of the program process.

But there is good reason to open the black box and investigate how the process works (or why it doesn't work). Consider recent research on welfare-to-work programs. The Manpower Demonstration Research Corporation reviewed findings from research on these programs in Florida, Minnesota, and Canada. In each location, adolescents with parents in a welfare-to-work program were compared with a control group of teenagers whose parents were on welfare but were not enrolled in welfare-to-work. In all three locations, teenagers in the welfare-to-work families actually did worse in school than those in the control group—troubling findings.

But why? Why did requiring welfare mothers to work hurt their children's schoolwork? Unfortunately, because the researchers had not investigated the program process—had not "opened the black box"—we can't know for sure. Martha Zaslow, an author of the resulting research report, speculated,

> Parents in the programs might have less time and energy to monitor their adolescents' behavior once they were employed. . . . Under the stress of working, they might adopt harsher parenting styles . . . the adolescents' assuming more responsibilities at home when parents got jobs was creating too great a burden. (cited in Lewin 2001b:A16)

But as Zaslow admitted, "We don't know exactly what's causing these effects, so it's really hard to say, at this point, what will be the long-term effects on these kids" (cited in Lewin 2001b:A16).

If an investigation of program process is conducted, a program theory may be developed. A program theory specifies how the program is expected to operate and identifies which program elements are operational (Chen 1990:32). It may identify the resources needed by the program and the activities that are essential to its operations, as well as how the program is to produce its effects (Fox et al. 2016:44). A program theory thus improves understanding of the relationship between the independent variable (the program) and the dependent variable (the outcome or outcomes). When a researcher has sufficient knowledge of this before the investigation begins, outlining a program theory can help guide the investigation of program process in the most productive directions. This is termed a theory-driven evaluation.

For example, Exhibit 13.9 illustrates the theory for an alcoholism treatment program. It shows that persons entering the program are expected to respond to the combination of motivational interviewing and peer support. A program theory can also decrease the risk of failure when the program is transported to other settings because it will help identify the conditions required for the program to have its intended effect.

Black box evaluation: The type of evaluation that occurs when an evaluation of program outcomes ignores, and does not identify, the process by which the program produced the effect.

Program theory: A descriptive or prescriptive model of how a program operates and produces effects.

Theory-driven evaluation: A program evaluation that is guided by a theory that specifies the process by which the program has an effect.

When it is possible to conduct multiple evaluation experiments, a program theory can be tested by repeating the impact analysis with different elements of the program omitted (Fox et al. 2016:96–97). For example, you can imagine repeating the impact analysis represented in Exhibit 13.9 after removing the peer support element of the treatment program, and then after removing the motivational interviewing element of the program.

Careers and Research

Mary Anne Casey, PhD, Consultant

Mary Anne Casey sailed through her undergraduate work without any exposure to social research. Her career in research and evaluation was never part of a “grand plan.” She just happened into it because of an assistantship in graduate school at the University of Minnesota. This graduate school experience—evaluating a regional foundation—fed her curiosity in research and evaluation.

After receiving her PhD, Casey worked for the State of Minnesota and the W. K. Kellogg Foundation and then joined a consulting firm. She weaves the lessons she has learned about research into her work, her writing on focus group interviewing (and a book with Richard Krueger on focus groups published by SAGE), and her teaching at the University of Minnesota, University of South Florida, and University of Michigan. Throughout her career, she has never stopped learning.

Each study is an opportunity to learn. I’ve learned about vexing issues and I’ve learned strategies that make me a better interviewer and analyst. The greatest reward is the honor of listening to people from a variety of backgrounds on intriguing topics: Midwest farmers on corn rootworms, veterans on their mental health care, mothers of new babies on home health care visits, teenagers on birth control, smokers on quitting, community members on garbage pickup, faculty on job satisfaction, and kids on what would get them to eat more fruits and vegetables. As a result, I know that there are multiple ways to see any issue. I believe this has made me less judgmental.

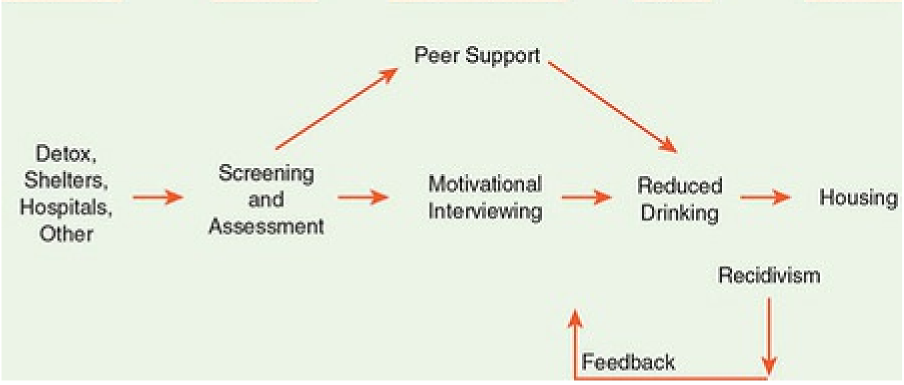
> Casey relishes analysis and finding just the right way to convey what people have shared. She urges students interested in research careers to hone their skills as listeners: “I hope my writing and teaching about focus group interviewing convinces others that careful listening is valuable and doable. We need good listeners.”

Program theory can be either descriptive or prescriptive (Chen 1990). Descriptive theory specifies what impacts are generated and how they occur. It suggests a causal mechanism, including intervening factors, and the necessary context for the effects. Descriptive theories are generally empirically based. Conversely, prescriptive theory specifies what the program ought to do but is not actually tested. Prescriptive theory specifies how to design or implement the treatment, what outcomes should be expected, and how performance should be judged. Comparison of the descriptive and prescriptive theories of the program can help identify implementation difficulties and incorrect understandings that can be corrected (Patton 2002:162–164).

Researcher or Stakeholder Orientation

Whose prescriptions specify how the program should operate, what outcomes it should try to achieve, or who it should serve? Most social science research assumes that the researcher specifies the research questions, the applicable theory or theories, and the outcomes to be investigated. Social science research results are most often reported in a professional journal or at professional conferences, where scientific standards determine how the research is received. In program evaluation, however, the research question is often set by the program sponsors or the government agency responsible for reviewing the program. In consulting projects for businesses, the client—a manager, perhaps, or a division president—decides what question researchers will study. Research findings are reported to these authorities. Most often, these authorities also specify the outcomes to be investigated. The first evaluator of the evaluation research is the funding agency, then, rather than the professional social science community. Evaluation research is research for a client, and its results may directly affect the services, treatments, or even punishments (e.g., in the case of prison studies) that program users receive. In this case, the person who pays the piper gets to call the tune.

Exhibit 13.9 The Program Theory for a Treatment Program for Homeless Alcoholics



Should the evaluation researcher insist on designing the evaluation project and specifying its goals, or should he or she accept the suggestions and adopt the goals of the funding agency? What role should the preferences of program staff or clients play? What responsibility does the evaluation researcher have to politicians and taxpayers when evaluating government-funded programs? The different answers that various evaluation researchers have given to these questions are reflected in different approaches to evaluation (Chen 1990:66–68).

Social science (or researcher) approaches emphasize the importance of researcher expertise and maintenance of some autonomy to develop the most trustworthy, unbiased program evaluation. It is assumed that "evaluators cannot passively accept the values and views of the other stakeholders" (Chen 1990:78). Evaluators who adopt this approach derive a program theory from information they obtain on how the program operates and extant social science theory and knowledge, not from the views of stakeholders. In one somewhat extreme form of this approach, *goal-free evaluation,* researchers do not even permit themselves to learn what goals the program stakeholders have for the program. Instead, the researcher assesses and then compares the needs of participants with a wide array of program outcomes (Scriven 1972b). The goal-free evaluator wants to see the unanticipated outcomes and to remove any biases caused by knowing the program goals in advance.

Stakeholder approaches encourage researchers to be responsive to program stakeholders (so this approach is also termed responsive evaluation). Issues for study are to be based on the views of people involved with the program, and reports are to be made to program participants (Shadish, Cook, and Leviton 1991:275–276). The program theory is developed by the researcher to clarify and develop the key stakeholders' theory of the program (Shadish et al. 1991:254–255). In one stakeholder approach, termed *utilization-focused evaluation,* the evaluator forms a task force of program stakeholders, who help to shape the evaluation project so that they are most likely to use its results (Patton 2002:171–175). In evaluation research termed *action research* or *participatory research* (discussed in Chapter 11), program participants are engaged with the researchers as coresearchers and help to design, conduct, and report the research. One research approach that has been termed *appreciative inquiry* eliminates the professional researcher altogether in favor of a structured dialogue about needed changes among program participants themselves (Patton 2002:177–185).

Social science (researcher) approach: An orientation to evaluation research that expects researchers to emphasize the importance of researcher expertise and maintenance of autonomy from program stakeholders.

Stakeholder approach: An orientation to evaluation research that expects researchers to be responsive primarily to the people involved with the program; also termed responsive evaluation.

In their book *Fourth Generation Evaluation,* Egon Guba and Yvonna Lincoln (1989) argue for evaluations oriented toward stakeholders:

> The stakeholders and others who may be drawn into the evaluation are welcomed as equal partners in every aspect of design, implementation, interpretation, and resulting action of an evaluation—that is, they are accorded a full measure of political parity and control . . . determining what questions are to be asked and what information is to be collected on the basis of stakeholder inputs. (p. 11)

Because different stakeholders may differ in their reports about or assessment of the program, there is not likely to be one conclusion about program impact. The evaluators are primarily concerned with helping participants understand the views of other stakeholders and with generating productive dialogue. Tineke Abma (2005) took this approach in a study of an injury prevention program at a dance school in the Netherlands:

> The evaluators acted as facilitators, paying deliberate attention to the development of trust and a respectful, open and comfortable climate. . . . Furthermore, the evaluation stimulated a public discourse about issues that were taboo, created a space for reflection, fostered dynamics and motivated participants to think about ways to improve the quality of their teaching practice. (pp. 284–285)

Of course, there are disadvantages in both stakeholder and social science approaches to program evaluation. If stakeholders are ignored, researchers may find that participants are uncooperative, that their reports are unused, and that the next project remains unfunded. However, if social science procedures are neglected, standards of evidence will be compromised, conclusions about program effects will likely be invalid, and results are unlikely to be generalizable to other settings. These equally undesirable possibilities have led to several attempts to develop more integrated approaches to evaluation research.

Integrative approaches attempt to cover issues of concern to both stakeholders and evaluators and to include stakeholders in the group from which guidance is routinely sought (Chen and Rossi 1987:101–102). The emphasis given to either stakeholder or social science concerns is expected to vary with the specific project circumstances. Integrated approaches seek to balance the goal of carrying out a project that is responsive to stakeholder concerns with the goal of objective, scientifically trustworthy, and generalizable results. When the research is planned, evaluators are expected to communicate and negotiate regularly with key stakeholders and to consider stakeholder concerns. Findings from preliminary inquiries are reported back to program decision makers so that they can make improvements in the program before it is formally evaluated. When the actual evaluation is conducted, the evaluation research team is expected to operate more autonomously, minimizing intrusions from program stakeholders.

Integrative approach: An orientation to evaluation research that expects researchers to respond to the concerns of people involved with the program—stakeholders—as well as to the standards and goals of the social scientific community.

Many evaluation researchers now recognize that they must account for multiple values in their research and be sensitive to the perspectives of different stakeholders, in addition to maintaining a commitment to the goals of measurement validity, internal validity, and generalizability (Chen 1990). Ultimately, evaluation research takes place in a political context, in which program stakeholders may be competing or collaborating to increase program funding or to emphasize particular program goals. A political process creates social programs, and a political process determines whether these programs are evaluated and what is done with the evaluation findings (Weiss 1993:94). Developing supportive relations with stakeholder groups will increase the odds that political processes will not undermine evaluation practice. You don't want to find out after you are finished that "people operating ineffective programs who depend on them for their jobs" are able to prevent an evaluation report from having any impact ("'Get Tough' Youth Programs" 2004:25).

Quantitative or Qualitative Methods

Evaluation research that attempts to identify the effects of a social program typically is quantitative: Did the response times of emergency personnel tend to decrease? Did the students' test scores increase? Did housing retention improve? Did substance abuse decline? It's fair to say that when there's an interest in comparing outcomes between an experimental and a control group or tracking change over time in a systematic manner, quantitative methods are favored.

But qualitative methods can add much to quantitative evaluation research studies, including more depth, detail, nuance, and exemplary case studies (Patton 2002). Perhaps the greatest contribution qualitative methods can make in many evaluation studies is investigating program process—finding out what is inside the black box. Although it is possible to track service delivery with quantitative measures such as frequency of staff contact and number of complaints, finding out what is happening to clients and how clients experience the program can often best be accomplished by observing program activities and interviewing staff and clients intensively.

For example, Michael Quinn Patton (2002:160) describes a study in which process analysis in an evaluation of a prenatal clinic's outreach program led to program changes. The process analysis revealed that the outreach workers were spending a lot of time responding to immediate problems, such as needs for rat control, protection from violence, and access to English classes. As a result, the outreach workers were recruiting fewer community residents for the prenatal clinic. New training and recruitment strategies were adopted to lessen this deviation from program goals.

Another good reason for using qualitative methods in evaluation research is the importance of learning how different individuals react to the treatment. For example, a quantitative evaluation of student reactions to an adult basic skills program for new immigrants relied heavily on the students' initial statements of their goals. However, qualitative interviews revealed that most new immigrants lacked sufficient experience in the United States to set meaningful goals; their initial goal statements simply reflected their eagerness to agree with their counselors' suggestions (Patton 2002:177–181).

Qualitative methods can also help reveal how social programs actually operate. Complex social programs have many different features, and it is not always clear whether the combination of those features or some particular features are responsible for the program's effect—or for the absence of an effect. Lisbeth Schorr, director of the Harvard Project on Effective Interventions, and Daniel Yankelovich, president of Public Agenda, put it this way: "Social programs are sprawling efforts with multiple components requiring constant midcourse corrections, the involvement of committed human beings, and flexible adaptation to local circumstances" (Schorr and Yankelovich 2000:A19).

The more complex the social program is, the more value qualitative methods can add to the evaluation process. Schorr and Yankelovich (2000) discuss the Ten Point Coalition, an alliance of black ministers that helped reduce gang warfare in Boston through multiple initiatives, “ranging from neighborhood probation patrols to safe havens for recreation” (p. A19). Qualitative methods would help describe a complex, multifaceted program such as this. A skilled qualitative researcher will be flexible and creative in choosing methods for program evaluation and will often develop mixed methods (described in Chapter 12), so that the evaluation benefits from the advantages of both qualitative and quantitative techniques.

Simple or Complex Outcomes

Does the program have only one outcome? Unlikely. How many outcomes are anticipated? How many might be unintended? Which are direct consequences of program action, and which are indirect effects that occur as a result of the direct effects (Mohr 1992)? Do the longer term outcomes follow directly from the immediate program outputs? Does the output (e.g., the increase in test scores at the end of the preparation course) result surely in the desired outcomes (i.e., increased rates of college admission)? Because of these and other possibilities, the selection of outcome measures is a critical step in evaluation research.

The decision to focus on one outcome rather than another, on a single outcome, or on several can have enormous implications. When Lawrence Sherman and Richard Berk (1984) evaluated the impact of an immediate arrest policy in cases of domestic violence in Minneapolis, they focused on recidivism as the key outcome. Similarly, the reduction of recidivism was the single desired outcome of prison "boot camps" opened in the 1990s. Boot camps are military-style programs for prison inmates that provide tough, highly regimented activities and harsh punishment for disciplinary infractions, with the goal of scaring inmates "straight." Boot camps were quite the rage in the 1990s, and the researchers who evaluated their impact understandably focused on criminal recidivism.

But these single-purpose programs turned out to be not quite so simple to evaluate. The Minneapolis researchers found that there was no adequate single source for records of recidivism in domestic violence cases, so they had to hunt for evidence from court and police records, follow-up interviews with victims, and family member reports (Sherman and Berk 1984). More easily measured variables, such as partners' ratings of the accused's subsequent behavior, eventually received more attention. Boot camp researchers soon concluded that the experience did not reduce recidivism: "Many communities are wasting a great deal of money on those types of programs" (Robert L. Johnson, cited in "'Get Tough' Youth Programs" 2004:25). However, some participants felt that the study had missed something (Latour 2002):

> [A staff member] saw things unfold that he had never witnessed among inmates and their caretakers. Those experiences profoundly affected the drill instructors and their charges, who still call to talk to the guards they once saw as torturers. Graduation ceremonies routinely reduced inmates, relatives, and sometimes even supervisors to tears. (p. B7)

A former boot camp superintendent, Michael Corsini, compared the Massachusetts boot camp with other correctional facilities and concluded, "Here, it was a totally different experience" (Latour 2002:B7).

Some observers now argue that the failure of boot camps to reduce recidivism results from the lack of postprison support rather than failure of the camps to promote positive change in inmates. Looking only at recidivism rates would be to ignore some important positive results.

Despite the additional difficulties introduced by measuring multiple outcomes, most evaluation researchers attempt to do so (Mohr 1992). When a program is evaluated initially only as a demonstration project, it is particularly important to consider the range of effects that might occur when it is adopted more widely in different communities (Fox et al. 2016:85). The result usually is a much more realistic, and richer, understanding of program impact.

Evaluation research on D.A.R.E. has also examined multiple outcomes. Most often, program impact is distinguished for type of drug use: alcohol, marijuana, and tobacco (Ennett et al. 1994:1399). Attitudinal change is also often examined. One common positive finding is of improved police–community relations and a more positive image of law enforcement in the eyes of students (Sloboda et al. 2009:7).

> It’s a very positive program for kids . . . a way for law enforcement to interact with children in a nonthreatening fashion . . . D.A.R.E. sponsored a basketball game. The middle school jazz band played. . . . We had families there. . . . D.A.R.E. officers lead activities at the [middle school]. . . . Kids do woodworking and produce a play. (Taylor 1999:1, 11)

For some people, this impact justifies maintaining the program even despite negative findings about its impact on drug abuse (Birkeland et al. 2005:248).

Measuring multiple outcomes may also lead to identification of different program impacts for different groups. In their evaluation of the alternative Take Charge of Your Life (TCYL) program for D.A.R.E. schools, for example, Sloboda and her colleagues (2009) found beneficial effects for students who reported marijuana use at baseline, but negative effects on alcohol and drug use for white students and concluded that the most effective approach to prevention may differ for those who have already had experience with illicit substances.

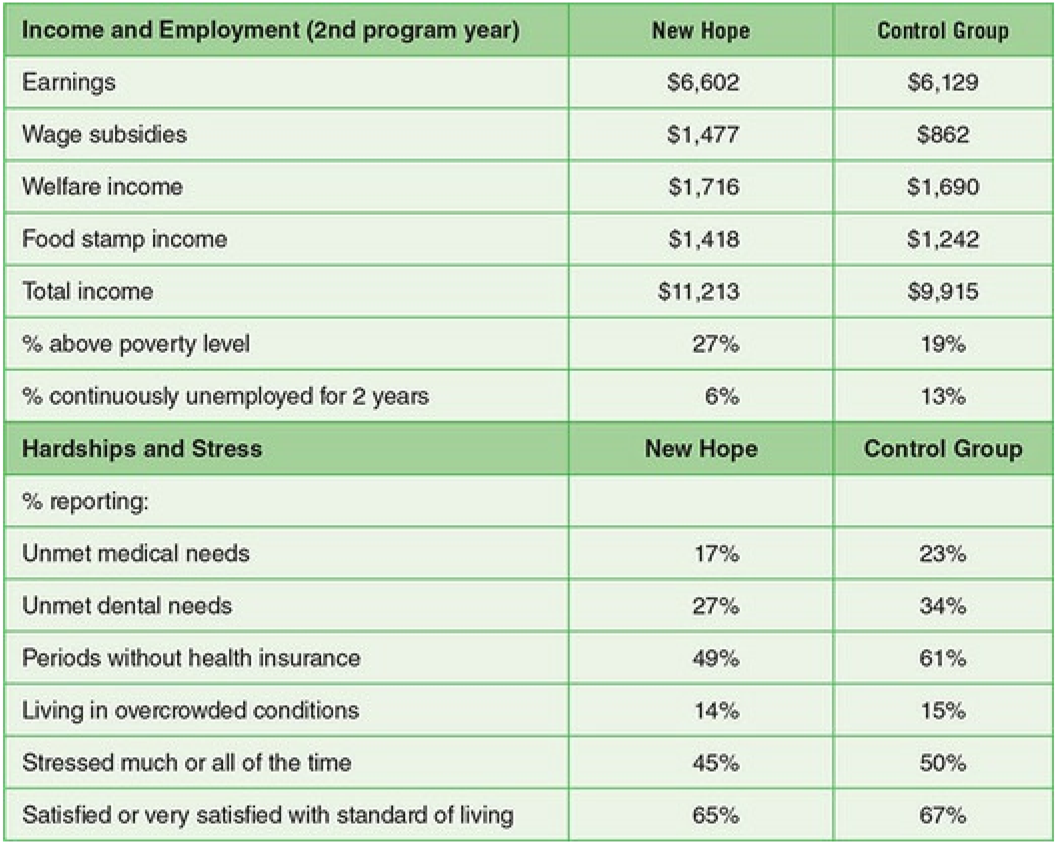
Project New Hope was an ambitious experimental evaluation of the impact of guaranteeing jobs to poor persons (DeParle 1999). It was designed to answer the following question: If low-income adults are given a job at a sufficient wage, above the poverty level, with child care and health care assured, how many would ultimately prosper? Some of the multiple outcomes measured in the evaluation of Project New Hope appear in Exhibit 13.10.

The project involved 677 low-income adults in Milwaukee who were offered a job involving work for 30 hours a week, along with child care and health care benefits. The outcome? Only 27% stuck with the job long enough to lift themselves out of poverty, and their earnings as a whole were only slightly higher than those of a control group that did not receive guaranteed jobs. Levels of depression were not decreased, or self-esteem increased, by the job guarantee. But there were some positive effects: The number of people who never worked at all declined, and rates of health insurance and use of formal child care increased. Perhaps most important, the classroom performance and educational hopes of participants' male children increased, with the boys' test scores rising by the equivalent of 100 points on the SAT and their teachers ranking them as better behaved.

So did the New Hope program "work"? Clearly it didn't live up to initial expectations, but it certainly showed that social interventions can have some benefits. Would the boys' gains continue through adolescence? Longer term outcomes would be needed. Why didn't girls (who were already performing better than the boys) benefit from their parents' enrollment in New Hope just as the boys did? A process analysis would have added a great deal to the evaluation design. The long and short of it is that collection of multiple outcomes gave a better picture of program impact.

Of course, there is a potential downside to the collection of multiple outcomes. Policy makers may choose to publicize only those outcomes that support their own policy preferences and ignore the rest. Often, evaluation researchers themselves have little ability to publicize a more complete story.

Exhibit 13.10 Outcomes in Project New Hope



Income and Employment (2nd program year)	New Hope	Control Group
Earnings	$6,602	$6,129
Wage subsidies	$1,477	$862
Welfare income	$1,716	$1,690
Food stamp income	$1,418	$1,242
Total income	$11,213	$9,915
% above poverty level	27%	19%
% continuously unemployed for 2 years	6%	13%
Hardships and Stress	**New Hope**	**Control Group**
% reporting:		
Unmet medical needs	17%	23%
Unmet dental needs	27%	34%
Periods without health insurance	49%	61%
Living in overcrowded conditions	14%	15%
Stressed much or all of the time	45%	50%
Satisfied or very satisfied with standard of living	65%	67%

Sources: DeParle (1999); Bos et al. (1999).

Groups or Individuals

Robert St. Pierre and Peter Rossi (2006) urge evaluation researchers to consider randomizing groups rather than individuals to alternative programs in an impact analysis. For example, a study of the effectiveness of a new educational program could randomly assign classes to either the new program or an alternative program rather than randomly assign individual students. Sloboda and her colleagues (2009) used this approach in their study of the alternative D.A.R.E. program, randomly assigning school districts to either the TCYL program or the control condition.

Randomization of groups to different treatments may be preferable when the goal is to compare alternative programs, so that different groups each receive some type of program. It can be easier to implement this approach if there are already different programs available and parents or other constituents are concerned that they (or their children) receive *some* type of program. Using group randomization also makes it easier to determine whether some characteristics of different sites (that offer different programs) influence program impact. However, this approach requires a larger number of participants and, often, cooperation across many governmental or organizational units (St. Pierre and Rossi 2006:667–675).

In a sense, all these choices (black box evaluation or program theory, researcher or stakeholder interests, and so on) hinge on (1) what your real goals are in doing the project and (2) how able you will be in a “research for hire” setting to achieve those goals. Not every agency really wants to know if its programs work, especially if the answer is no. Dealing with such issues, and the choices they require, is part of what makes evaluation research both scientifically and politically fascinating.

Policy Research

Policy research is a process rather than a method: "a process that attempts to support and persuade actors by providing them with well-reasoned, evidence-based, and responsible recommendations for decision making and action" (Majchrzak and Markus 2014:3). Because policy research often draws on the findings of evaluation research projects and involves working for a client, as is the case in evaluation research, policy researchers confront many of the same challenges as do evaluation researchers. Because policy researchers must summarize and weigh evidence from a wide range of sources, they need to be familiar with each of the methods presented in this book.

The goal of policy research is to inform those who make policy about the possible alternative courses of action in response to some identified problem, their strengths and weaknesses, and their likely positive and negative effects. Reviewing the available evidence may lead the policy researcher to conclude that enough is known about the issues to develop recommendations without further research, but it is more likely that additional research will be needed using primary or secondary sources. Policy researchers Ann Majchrzak and M. Lynne Markus (2014:9) caution those embarking on a policy research project to ensure that their plans are

- *Credible:* informed by evidence and unbiased to the extent possible about the pros, cons, and risk of problems and potential interventions
- *Meaningful and engaging* of representatives of stakeholder groups, including policy makers and regulators, and those affected by policy actions such as customers, suppliers, or service recipients
- *Responsible:* consider a broad spectrum of potential negative consequences of policy change
- *Creative:* recognize needs for new or different solutions
- *Manageable:* doable within the available time and resources

The policy research process begins with identification of a research question that may focus on the causes of or solutions to the policy problem. For example, Peter Reuter, Rosalie Liccardo Pacula, and Jonathan Caulkins (2010) describe one of the policy research questions on which RAND's Drug Policy Research Center (DPRC) has focused as Getting to Outcomes. In this project, prevention scholars, community organizers, and service researchers in the DPRC reviewed the literature and then developed manuals, worksheets, and other resources to help communities adopting prevention programs.

Policy research: A process in which research results are used to provide policy actors with recommendations for action that are based on empirical evidence and careful reasoning.

Majchrzak and Markus (2014:22–25) encourage policy researchers to make a clear distinction between the problem they hope to solve and the aspects of the problem they cannot deal with, to review carefully the context in which the problem occurs, to specify clearly why policy change is needed and what possible risks might be incurred by making a change, as well as to develop a causal model of how the policy problem occurs. RAND policy researchers (Reuter et al. 2010:257) found that one of the greatest challenges for designing quality policy research is the longer time frame required for producing a good product that involves primary data collection, compared with the short time frame required with most politically driven policy decisions. Stefanie Ettelt, Nicholas Mays, and Ellen Nolte (2013) register a similar caution based on their experience in a policy research center funded by the British Department of Health to produce rapid analyses of health care experiences in other countries. Nonetheless, the development of centers like these suggests that policy research is becoming an expected element in policy making. These problems in translating evaluation research results into practice explain why many researchers feel that the best they can hope for is policy “informed” by research evidence, rather than “based” on evidence (Fox et al. 2016:13).

Ethics in Evaluation

Evaluation research can make a difference in people's lives while the research is being conducted as well as after the results are reported. Job opportunities, welfare requirements, housing options, treatment for substance abuse, and training programs are each potentially important benefits, and an evaluation research project can change both the type and the availability of such benefits. This direct impact on research participants and, potentially, their families, heightens the attention that evaluation researchers have to give to human subject concerns (Wolf, Turner, and Toms 2009:171). Although the particular criteria that are at issue and the decisions that are judged most ethical vary with the type of evaluation research conducted and the specifics of a particular project, there are always serious ethical as well as political concerns for the evaluation researcher (Boruch 1997:13; Dentler 2002:166).

When program impact is the focus, human subject considerations multiply. What about assigning persons randomly to receive some social program or benefit? One justification for this given by evaluation researchers has to do with the scarcity of these resources. If not everyone in the population who is eligible for a program can receive it, because of resource limitations, what could be a fairer way to distribute the program benefits than through a lottery? Random assignment also seems like a reasonable way to allocate potential program benefits when a new program is being tested with only some members of the target recipient population. If the value of a program is not really known, random assignment to an alternative version seems particularly reasonable (Fox et al. 2016:36–37). However, when an ongoing entitlement program is being evaluated and experimental subjects would normally be eligible for program participation, it may not be ethical simply to bar some potential participants from the programs. Instead, evaluation researchers may test alternative treatments or provide some alternative benefit while the treatment is being denied.

There are many other ethical challenges in evaluation research:

- How can confidentiality be preserved when the data are owned by a government agency or are subject to discovery in a legal proceeding?
- Who decides what level of burden an evaluation project may tolerably impose on participants?
- Is it legitimate for research decisions to be shaped by political considerations?
- Must evaluation findings be shared with stakeholders rather than only with policy makers?
- Is the effectiveness of the proposed program improvements really uncertain?
- Will a randomized experiment yield more defensible evidence than the alternatives?
- Will the results actually be used?

The Health Research Extension Act of 1985 (Public Law 99–158) mandated that the Department of Health and Human Services require all research organizations receiving federal funds to have an institutional review board to assess all research for adherence to ethical practice guidelines. We have already reviewed the federally mandated criteria (Boruch 1997:29–33):

- Are risks minimized?
- Are risks reasonable in relation to benefits?
- Is the selection of individuals equitable? (randomization implies this)
- Is informed consent given?
- Are the data monitored?
- Are privacy and confidentiality assured?

Evaluation researchers must consider whether it will be possible to meet each of these criteria long before they even design a study. It is important to inform representatives of all stakeholders about the evaluation before it begins and to consider with them the advantages and disadvantages for potential participants and the type of protections needed (Fox et al. 2016:28–31).

The problem of maintaining subject confidentiality is particularly thorny because researchers, in general, are not legally protected from the requirements that they provide evidence requested in legal proceedings, particularly through the process known as "discovery." However, it is important to be aware that several federal statutes have been passed specifically to protect research data about vulnerable populations from legal disclosure requirements. For example, the Crime Control and Safe Streets Act (28 CFR Part 11) includes the following stipulation:

> Copies of [research] information [about persons receiving services under the act or the subjects of inquiries into criminal behavior] shall be immune from legal process and shall not, without the consent of the persons furnishing such information, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceedings. (Boruch 1997:60)

Ethical concerns must also be given special attention when evaluation research projects involve members of vulnerable populations as subjects. To conduct research on children, parental consent usually is required before the child can be approached directly about the research. Adding this requirement to an evaluation research project can dramatically reduce participation because many parents simply do not bother to respond to mailed consent forms. Sloboda and colleagues' (2009:3) evaluation of the trial TCYL program used an "active consent" procedure for gaining parental consent and student assent: parents and students both had to sign forms before the student could participate; the result was that

58% of the 34,000 eligible seventh-grade students were enrolled in the study. Other research indicates that use of a “passive consent” procedure—students can participate as long as their parents do not return a form indicating their lack of consent—can result in much higher rates of participation. Since nonconsent is likely to be higher among those who are more at risk of substance abuse, the likelihood of identifying program impact can be diminished (Tigges 2003).

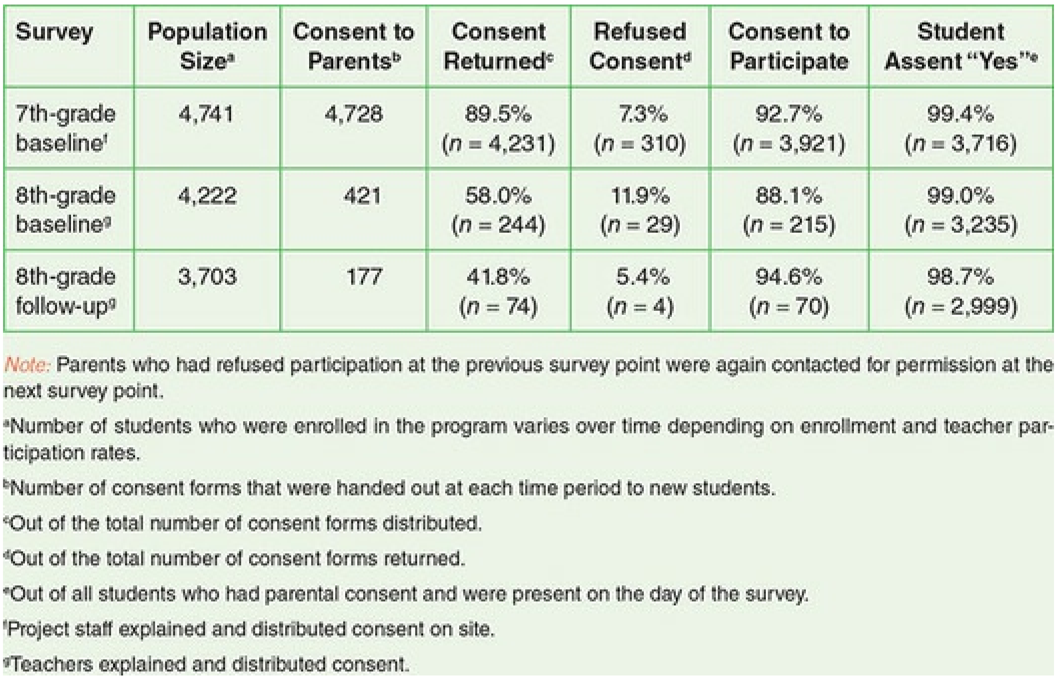
Tricia Leakey and her colleagues (2004:511) demonstrated that this problem can be overcome in their evaluation of Project SPLASH (Smoking Prevention Launch Among Students in Hawaii). When the project began in the seventh grade, the researchers gave students project information and a consent card to take home to their parents. A pizza party was then held in every class where at least 90% of the students returned a signed consent card. In subsequent follow-ups in the eighth grade, a reminder letter was sent to parents whose children had previously participated. Classes with high participation rates also received a candy thank you. As you can see in Exhibit 13.11, the result was a very high rate of participation.

When it appears that it will be difficult to meet the ethical standards in an evaluation project, at least from the perspective of some of the relevant stakeholders, modifications should be considered in the study design. Several steps can be taken to lessen any possibly detrimental program impact (Boruch 1997:67–68):

- Alter the group allocation ratios to minimize the number in the untreated control group.
- Use the minimum sample size required to be able to adequately test the results.
- Test just parts of new programs rather than the entire program.
- Compare treatments that vary in intensity (rather than presence or absence).
- Vary treatments between settings rather than between individuals within a setting.

Essentially, each of these approaches limits the program’s impact during the experiment and so lessens any potential adverse effects on human subjects. It is also important to realize that it is costly to society and potentially harmful to participants to maintain ineffective programs. In the long run, at least, it may be more ethical to conduct an evaluation study in order to improve a program than to let the status quo remain in place (Fox et al. 2016:94).

Exhibit 13.11 Parental Consent Response Rates and Outcomes



Survey	Population Size[a]	Consent to Parents[b]	Consent Returned[c]	Refused Consent[d]	Consent to Participate	Student Assent "Yes"[e]
7th-grade baseline[f]	4,741	4,728	89.5% (*n* = 4,231)	7.3% (*n* = 310)	92.7% (*n* = 3,921)	99.4% (*n* = 3,716)
8th-grade baseline[g]	4,222	421	58.0% (*n* = 244)	11.9% (*n* = 29)	88.1% (*n* = 215)	99.0% (*n* = 3,235)
8th-grade follow-up[g]	3,703	177	41.8% (*n* = 74)	5.4% (*n* = 4)	94.6% (*n* = 70)	98.7% (*n* = 2,999)

Note: Parents who had refused participation at the previous survey point were again contacted for permission at the next survey point.

[a] Number of students who were enrolled in the program varies over time depending on enrollment and teacher participation rates.

[b] Number of consent forms that were handed out at each time period to new students.

[c] Out of the total number of consent forms distributed.

[d] Out of the total number of consent forms returned.

[e] Out of all students who had parental consent and were present on the day of the survey.

[f] Project staff explained and distributed consent on site.

[g] Teachers explained and distributed consent.

Source: Leakey, Tricia, Kevin B. Lunde, Karin Koga, and Karen Glanz. 2004. "Written Parental Consent and the Use of Incentives in a Youth Smoking Prevention Trial: A Case Study From Project SPLASH." American Journal of Evaluation 25:509–523.

Conclusions

Hopes for evaluation research are high: Society could benefit from the development of programs that work well, that accomplish their policy goals, and that serve the people who genuinely need them. At least that is the hope. Unfortunately, there are many obstacles to realizing this hope (Posavac and Carey 1997):

- Because social programs and the people who use them are complex, evaluation research designs can easily miss important outcomes or aspects of the program process.
- Because the many program stakeholders all have an interest in particular results from the evaluation, researchers can be subjected to an unusual level of cross-pressures and demands.
- Because the need to include program stakeholders in research decisions may undermine adherence to scientific standards, research designs can be weakened.
- Because some program administrators want to believe that their programs really work well, researchers may be pressured to avoid null findings or, if they are not responsive, may find their research report ignored. Plenty of well-done evaluation research studies wind up in a recycling bin or hidden away in a file cabinet.
- Because the primary audience for evaluation research reports are program administrators, politicians, or members of the public, evaluation findings may need to be overly simplified, distorting the findings.

The rewards of evaluation research are often worth the risks, however. Evaluation research can provide social scientists with rare opportunities to study a complex social process, with real consequences, and to contribute to the public good. Although they may face unusual constraints on their research designs, most evaluation projects can result in high-quality analysis and publications in reputable social science journals. In many respects, evaluation research is an idea whose time has come. We may never achieve Donald Campbell's (Campbell and Russo 1999) vision of an "experimenting society," in which research is consistently used to evaluate new programs and to suggest constructive changes, but we are close enough to continue trying.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Black box evaluation	505
Cost–benefit analysis	503
Cost–effectiveness analysis	503
Efficiency analysis	503
Evaluability assessment	495
Feedback	491
Formative evaluation	498
Impact evaluation (or analysis)	499
Inputs	491
Integrative approach	508
Needs assessment	492
Outcomes	491
Outputs	491
Policy research	513
Process evaluation	496
Program process	491
Program theory	505
Responsive evaluation	507
Social science (researcher) approach	507
Stakeholder approach	507
Stakeholders	491
Summative evaluation	499
Theory-driven evaluation	505

Highlights

- Evaluation research is social research that is conducted for a distinct purpose: to investigate social programs.
- The development of evaluation research as a major enterprise followed on the heels of the expansion of the federal government during the Great Depression and World War II.
- The evaluation process can be modeled as a feedback system, with inputs entering the program, which generates outputs and then outcomes, which feed back to program stakeholders and affect program inputs.
- The evaluation process as a whole, and the feedback process in particular, can be understood only in relation to the interests and perspectives of program stakeholders.
- There are five primary types of program evaluation: (1) needs assessment, (2) evaluability assessment, (3) process evaluation (including formative evaluation), (4) impact evaluation (also known as summative evaluation), and (5) efficiency (cost–benefit) analysis.
- The process by which a program has an effect on outcomes is often treated as a "black box," but there is good reason to open the black box and investigate the process by which the program operates and produces, or fails to produce, an effect.
- A program theory may be developed before or after an investigation of the program process is completed. It may be either descriptive or prescriptive.

- Evaluation research is done for a client, and its results may directly affect the services, treatments, or punishments that program users receive. Evaluation researchers differ in the extent to which they attempt to orient their evaluations to program stakeholders.
- Qualitative methods are useful in describing the process of program delivery.
- Multiple outcomes are often necessary to understand program effects.
- Policy research is a process, rather than a specific method, that uses "well-reasoned, evidence-based, responsible recommendations" to inform policy decisions.
- Evaluation research raises complex ethical issues because it may involve withholding desired social benefits.

Discussion Questions

1. Would you prefer that evaluation researchers use a stakeholder or a social science approach? Compare and contrast these perspectives, and list at least four arguments for the one you favor.
2. Propose a randomized experimental evaluation of a social, medical, or educational program with which you are familiar. Possibilities could range from a job training program to a community health center, or even a college. Include in your proposal a description of the program and its intended outcomes. Discuss the strengths and weaknesses of your proposed design.
3. How would you describe the contents of the "black box" of program operations? What "program theory" would specify how the program (in Question 2) operates?
4. What would be the advantages and disadvantages of using qualitative methods to evaluate this program? What would be the advantages and disadvantages of using quantitative methods? Which approach would you prefer and why?

Practice Exercises

1. Read and summarize the evaluation research by Amie Schuck that begins this chapter. Be sure to identify the type of evaluation research that is described. Discuss the strengths and weaknesses of the design.
2. Identify the key stakeholders in a local social or educational program. Interview several stakeholders to determine what their goals for the program are and what tools they use to assess goal achievement. Compare and contrast the views of each stakeholder and try to account for any differences you find.
3. Review the “Evaluation Research” lesson in the interactive exercises on the book’s study site, at edge.sagepub.com/schutt9e, to learn more about the language and logic of evaluation research.
4. Identify an article that reports an evaluation research study on the book’s study site. What type of evaluation research does this study represent? What alternatives did the author(s) select when designing the research? After reading the entire article, do you agree with the author’s (or authors’) choices? Why or why not?

Ethics Questions

1. Imagine that you are evaluating a group home for persons with serious mental illness and learn that a house resident has been talking about cutting himself. Would you immediately inform house staff about this? What if the resident asked you not to tell anyone? In what circumstances would you feel it is ethical to take action to prevent the likelihood of a subject harming himself or herself or others?
2. Is it ethical to assign people to receive some social benefit on a random basis? Form two teams, and debate the ethics of the TARP randomized evaluation of welfare payments described in this chapter.

Web Exercises

1. Inspect the website maintained by the Governmental Accounting Standards Board (GASB), at www.gasb.org, and particularly the section at http://www.gasb.org/jsp/GASB/Page/GASBSectionPage&cid=1176156714545. Describe the GASB process of standard-setting.
2. Describe the resources available for evaluation researchers at one of the following three websites: The Evaluation Center at Western Michigan University (www.wmich.edu/evaluation), the National Network of Libraries of Medicine (https://nnlm.gov/mcr/professional-development/program-evaluation/evaluation-sites-web), and the Independent Evaluation Group at the World Bank Group (http://ieg.worldbankgroup.org/).
3. You can check out the latest information regarding the D.A.R.E. program at www.dare.com. Check out "About" "Mission|Vision" (https://www.dare.org/mission-vision/). What is the current approach? Can you find information on the web about current research on D.A.R.E.?
4. Evaluation research is a big industry! Two examples are provided by Mathematica Policy Research (https://www.mathematica-mpr.com) and the Policy Evaluation and Research Center at Educational Testing Services (www.ets.org/research/perc/). Summarize their work.

Video Interview Questions

Listen to the researcher interview for Chapter 13 at edge.sagepub.com/schutt9e.

1. Why was this specific research study challenging?
2. How did the researchers come up with the “counterfactual” component of the study?

SPSS Exercises

1. Neighborhood and school integration has often been a focus of government social policy. Does the racial composition of a neighborhood have any association with attitudes related to racial issues? Although we cannot examine the effects of social policies or programs directly in the General Social Survey (GSS) data, we can consider the association between neighborhood racial composition and attitudes related to race. The variable RACLIVE indicates whether the respondent lives in a racially integrated neighborhood. Request its frequency distribution as well as those for several attitudes related to race: RACOPEN, AFFRMACT, WRKWAYUP, HELPBLK, and CLOSEBLK3.
2. Do attitudes vary with the experience of living in a racially integrated neighborhood? Request the cross-tabulation of the variables used in Step 1, RACOPEN to CLOSEBLK3 by RACLIVE (request percentages on the column totals). Read the tables and explain what they tell us about attitudes and neighborhoods. Does the apparent effect of racial integration vary with the different attitudes? How would you explain this variation in these “multiple outcomes”?
3. What other attitudes differ between whites who live in integrated and segregated neighborhoods? Review the GSS2016 or GSS2016x variable list to identify some possibilities and request cross-tabulations for these variables. Do you think these differences are more likely to be a consequence of a racially integrated neighborhood experience or a cause of the type of neighborhood that people choose to live in? Explain.

Developing a Research Proposal

If you plan an evaluation research project, you will have to revisit the decisions about research designs (Exhibit 3.9, #13 to #17).

1. Develop a brief model for a program that might influence the type of attitude or behavior in which you are interested. List the key components of this model.
2. Design a program evaluation to test the efficacy of your program model, using an impact analysis approach.
3. Add to your plan a discussion of a program theory for your model. In your methodological plan, indicate whether you will use qualitative or quantitative techniques and simple or complex outcomes.
4. Who are the potential stakeholders for your program? How will you relate to them before, during, and after your evaluation?
5. What steps would you include in a policy research project designed to develop recommendations for policy makers about this type of attitude or behavior?

Chapter 14 Research Using Secondary Data and "Big" Data

- Secondary Data Sources
- Research That Matters, Questions That Count
- Careers and Research
    - U.S. Census Bureau
    - Integrated Public Use Microdata Series
    - Bureau of Labor Statistics
    - Other Government Sources
    - Other Data Sources
    - Inter-university Consortium for Political and Social Research
        - *Types of Data Available From ICPSR*
        - *Obtaining Data From ICPSR*
    - Harvard's Dataverse
    - International Data Sources
    - Qualitative Data Sources
- Challenges for Secondary Data Analyses
- Big Data
    - Background
    - Examples of Research Using Big Data
- Ethical Issues in Secondary Data Analysis and Big Data
- Research in the News: A Bright Side to Facebook's Experiments on Its Users?
- Conclusions

Irish researchers Richard Layte (Economic and Social Research Institute) and Christopher Whelan (University College Dublin) sought to improve understanding of poverty in Europe. Rather than design their own data collection effort, they turned to five waves of data from the European Community Household Panel survey, which were available to them from Eurostat, the Statistical Office of the European Communities (Eurostat 2003). The data they obtained represented the years from 1994 to 1998, thus allowing Layte and Whelan (2003) to investigate whether poverty tends to persist more in some countries than in others and what factors influence this persistence in different countries. Their investigation of "poverty dynamics" found a tendency for individuals and households to be "trapped" in poverty, but this phenomenon varied with the extent to which countries provided social welfare supports. Like the analysis of income inequality and household debt in the United States by Fligstein, Hastings, and Goldstein, the local context mattered.

Secondary data analysis is the method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data. The most common sources of secondary data—previously collected data that are used in a new analysis—are social science surveys and data collected by government agencies, often with survey research methods. It is also possible to reanalyze data that have been collected in experimental studies or with qualitative methods. Even a researcher's reanalysis of data that he or she collected previously qualifies as secondary analysis if it is employed for a new purpose or in response to a methodological critique.

Thanks to the data collected by social researchers, governments, and organizations over many years, secondary data analysis has become the research method used by many contemporary social scientists to investigate important research questions. Why consider secondary data? (1) Data collected in previous investigations are available for use by other social researchers on a wide range of topics. (2) Available data sets often include many more measures and cases and reflect more rigorous research procedures than another researcher will have the time or resources to obtain in a new investigation. (3) Much of the groundwork involved in creating and testing measures with the data set has already been done. (4) Most important, most funded social science research projects collect data that can be used to investigate new research questions that the primary researchers who collected the data did not consider. Combining data sets in the way that Fligstein, Hastings, and Goldstein (2017) did allows even more innovative investigations. Analyzing secondary data, then, is nothing like buying "used goods"!

I will first review the procedures involved in secondary data analysis, identify many of the sources for secondary data sets, and explain how to obtain data from these sources. I will give special attention to some easy-to-overlook problems with the use of secondary data. I then introduce the concept of "Big Data," which involves the analysis of very large secondary data sets that often contain digital records of the behavior of many individuals. The chapter concludes with some ethical cautions related to the use of these methods.

Secondary data analysis: The method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data.

Secondary data: Previously collected data that are used in a new analysis.

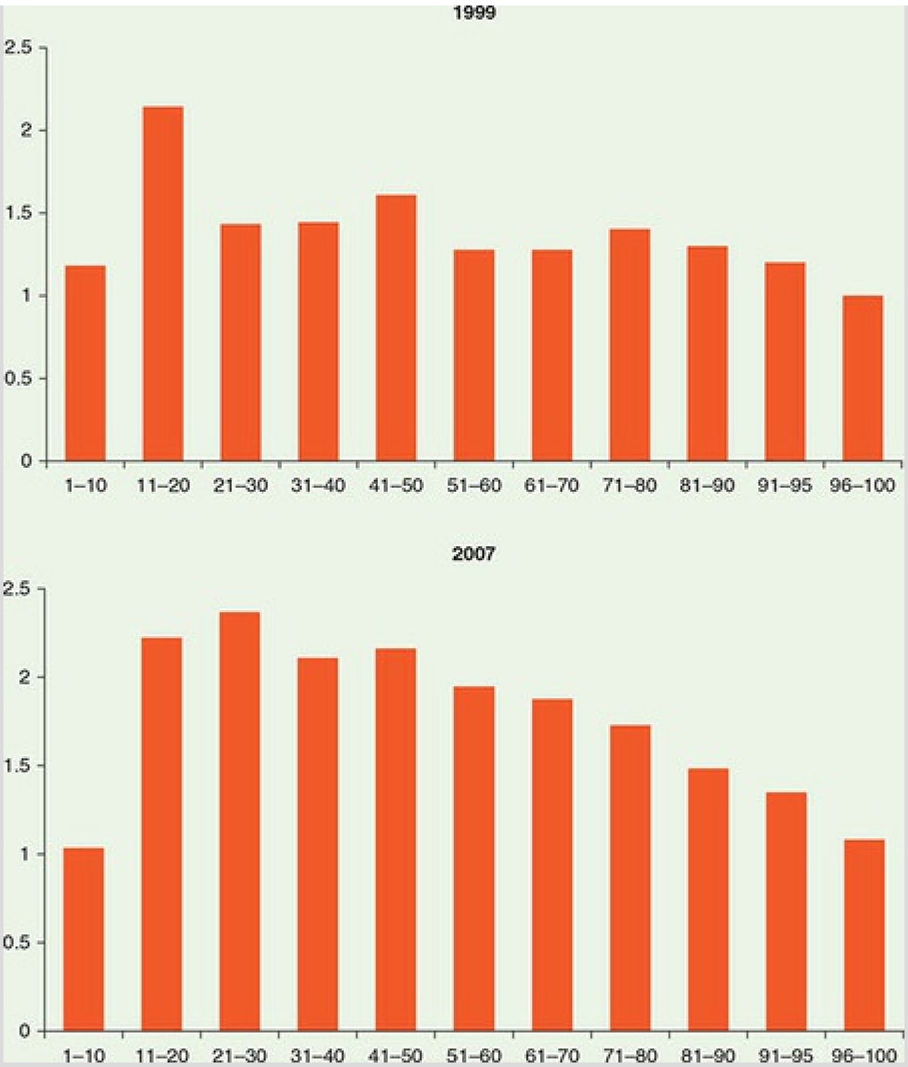
Secondary Data Sources

Secondary data analysis has been an important social science methodology since the earliest days of social research, whether when Karl Marx (1967) reviewed government statistics in the Reading Room of the British Library during the 1850s to 1870s or when Émile Durkheim (1966) analyzed official government cause-of-death data for his study of suicide rates throughout Europe in the late 19th century. With the advent of modern computers and, even more important, the Internet, secondary data analysis has become an increasingly accessible social research method. Literally thousands of large-scale data sets are now available for the secondary data analyst, often with no more effort than the few commands required to download the data set. A number of important data sets can even be analyzed directly on the web by users who lack their own statistical software.

Research That Matters, Questions That Count

Growing income inequality in the United States has made it hard for many households to maintain the lifestyle to which they are accustomed. In numerous urban areas, high housing prices are part of the problem, particularly for those who move and seek a new home that lives up to their expectations. Sociologists Neil Fligstein, Orestes P. Hastings, and Adam Goldstein (2017) decided to investigate this problem to learn about the dynamic relationship between inequality, lifestyle, and consumer consumption.

Exhibit 14.1 Increasing Mean Debt-to-Income Ratio by Income Group, 1999, 2007



Source: Fligstein et al. (2017:8).

Identifying status competition—“keeping up with Joneses”—as a potential key influence in this relationship, Fligstein, Hastings, and Goldstein hypothesized that in neighborhoods with more income inequality, homeowners would tend to spend more to buy homes and take on more debt to “keep up.” Their analysis combined data from the Panel Study of Income Dynamics (PSID) that tracks consumer expenditures with data from the housing industry’s Zillow database of housing prices and Internal Revenue Service data about income inequality. They focused attention on 4,354 households that moved between 1999 and 2007. You can see in Exhibit 14.1 that the mean (average) debt-to-income ratio for homeowners during this period got much worse for most income groups, compared to those in the highest income groups.

The analysis revealed that “increasing income inequality allowed the highest income households to move to the best neighborhoods while pushing everyone in unequal areas to spend more money to maintain their lifestyles” (Fligstein et al. 2017:13).

1. What evidence have you seen that people do their best to maintain their lifestyle relative to their neighbors, even in the face of diminished income?
2. What might be missing from an analysis that focuses only on the available data about inequality, expenditures, and housing costs? What questions would you like to include in a survey of homebuyers to better understand their behavior?

In this chapter, you will learn the basic logic and procedures that guide secondary data and Big Data analysis. By the end of the chapter, you will understand the appeal of these research approaches and be able to identify their strengths, limitations, and many of the sources of such data. As you complete the chapter, you can enrich your understanding by reading the 2017 *Socius* article by Fligstein, Hastings, and Goldstein at the *Investigating the Social World* study site and by completing the interactive exercises for Chapter 14 at edge.sagepub.com/schutt9e.

Source: Fligstein, Neil, Orestes P. Hastings, and Adam Goldstein. 2017. “Keeping Up With the Joneses: How Households Fared in the Era of High Income Inequality and the Housing Price Bubble, 1999–2007.” *Socius: Sociological Research for a Dynamic World* 3:1–15.

There are many sources of data for secondary analysis within the United States and internationally. These sources range from data compiled by governmental units and private organizations for administrative purposes, which are subsequently made available for research purposes, to data collected by social researchers for one purpose that are then made available for reanalysis. Many important data sets are collected for the specific purpose of facilitating secondary data analysis. Government units from the Census Bureau to the U.S. Department of Housing and Urban Development; international organizations such as the United Nations, the Organisation for Economic Co-operation and Development (OECD), and the World Bank; and internationally involved organizations such as the Central Intelligence Agency (CIA) sponsor a substantial amount of social research that is intended for use by a broader community of social scientists. The National Opinion Research Center (NORC), with its General Social Survey (GSS), and the University of Michigan, with its Detroit Area Studies, are examples of academically based research efforts that are intended to gather data for social scientists to use in analyzing a range of social science research questions.

Many social scientists who have received funding to study one research question have subsequently made the data they collect available to the broader social science community for investigations of other research questions. Many of these data sets are available from a website maintained by the original research organization, often with some access restrictions. Examples include the Add Health study conducted at the University of North Carolina Population Center, the University of Michigan’s Health and Retirement Study as well as its Detroit Area Studies, and the United Nations University’s World Income Inequality Database.

What makes secondary data analysis such an exciting and growing option today are the considerable resources being devoted to expanding the amount of secondary data and to making it available to social scientists. For example, the National Data Program for the Social Sciences, funded in part by the National Science Foundation, sponsors the ongoing GSS to make current data on a wide range of research questions available to social scientists. Since 1985, the GSS has participated in an International Social Survey Program that generates comparable data from 47 countries around the world (www.issp.org). Another key initiative is the Data Preservation Alliance for the Social Sciences (Data-PASS), funded by the Library of Congress in 2004 as a part of the National Digital

Preservation Program
(http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/inde
This project is designed to ensure the preservation of digitized social science data. Led by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan, it combines the efforts of other major social research organizations, including the Roper Center for Public Opinion Research at the University of Connecticut; the Howard W. Odum Institute for Research in Social Sciences at the University of North Carolina, Chapel Hill; the Henry A. Murray Research Archive and the Harvard-MIT Data Center at Harvard University; and the Electronic and Special Media Records Service Division of the U.S. National Archives and Records Administration.

Fortunately, you do not have to google your way around the web to find all these sources on your own. Many websites provide extensive collections of secondary data. Chief among these is the ICPSR website at the University of Michigan. The University of California at Berkeley's Survey Documentation and Analysis (SDA) archive provides several data sets from national omnibus surveys, as well as from U.S. Census microdata, from surveys on racial attitudes and prejudice, and from several labor and health surveys. The National Archive of Criminal Justice Data is an excellent source of data in the area of criminal justice, although, like many other data collections, including key data from the U.S. Census, it is also available through ICPSR. Much of the statistical data collected by U.S. federal government agencies can be accessed through the consolidated FedStats website, https://fedstats.sites.usa.gov.

In this section, I will describe several sources of online data in more detail. The decennial population census by the U.S. Census Bureau is the single most important governmental data source, but many other data sets are collected by the U.S. Census and by other government agencies, including the U.S. Census Bureau's *Current Population Survey (CPS)* and its Survey of Manufactures or the Bureau of Labor Statistics' Consumer Expenditure Survey. These government data sets typically are quantitative; in fact, the term *statistics*—state-istics—is derived from this type of data.

Inter-university Consortium for Political and Social Research (ICPSR): The academic consortium that archives data sets online from major surveys and other social science research and makes them available for analysis by others.

Careers and Research

Lee Rainie, Pew Research Center

You already know of Lee Rainie's research from findings presented in Chapter 1 from the Pew Research Center Internet Project. Rainie is a graduate of Harvard University and has an MA in political science from Long Island University. He was for many years managing editor at *U.S. News & World Report,* but since 1999, he has directed the Pew Internet Project, a nonprofit, nonpartisan "fact tank" that studies the social impact of the Internet. Since December 1999, the Washington, D.C., research center has explored how people's Internet use affects their families, communities, health care, education, civic and political life, and workplaces.

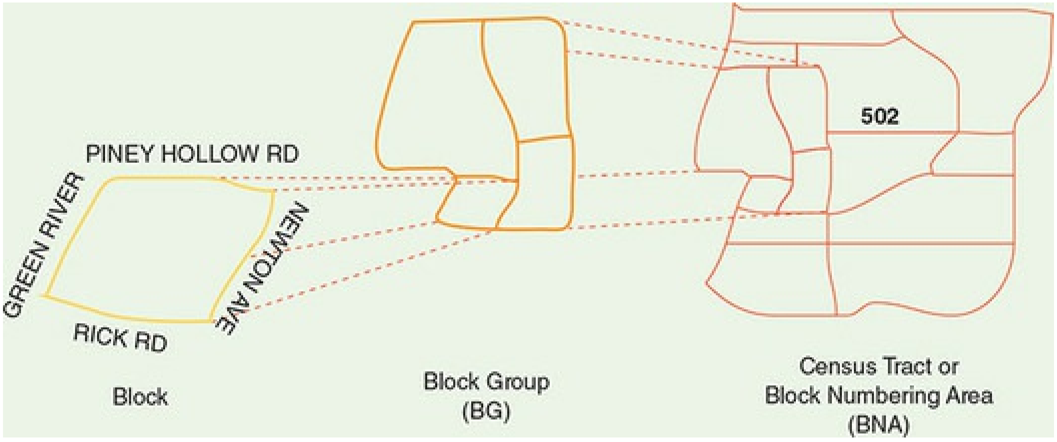
The project is funded by the Pew Charitable Trusts and has issued more than 500 reports based on its surveys that examine people's online activities and the Internet's role in their lives. All of its reports and data sets are available online for free at www.pewinternet.org. The value of their work is apparent in its wide public impact. Rainie and other project staff have testified before Congress on the new media environment, privacy, and family issues related to Internet use. They have also given briefings and presentations to White House officials; several government commissions; the Federal Communications Commission; the Federal Trade Commission; the U.S. Departments of Commerce, Health and Human Services, and Agriculture; the U.S. Conference of Governors, the National Institutes of Health; the Centers for Disease Control and Prevention; the National Conference of State Legislators; and hundreds of other local, state, and federal officials. Project findings are used by the U.S. Census Bureau, the Organisation for Economic Co-operation and Development (OECD), and the World Economic Forum communications and media group. Many researchers use data collected by the Pew Internet Project as the foundation for secondary data analysis projects.

U.S. Census Bureau

The U.S. government has conducted a census of the population every 10 years since 1790; since 1940, this census has also included a census of housing (see Chapter 5). This decennial Census of Population and Housing is a rich source of social science data (Lavin 1994). The Census Bureau's monthly *CPS* provides basic data on labor force activity that is then used in U.S. Bureau of Labor Statistics reports. The Census Bureau also collects data on agriculture, manufacturers, construction and other business, foreign countries, and foreign trade.

The U.S. Census of Population and Housing aims to survey one adult in every household in the United States. The basic *complete-count* census contains questions about household composition as well as ethnicity and income. More questions are asked in a longer form of the census that is administered to a sample of the households. A separate census of housing characteristics is conducted at the same time (Rives and Serow 1988:15). Participation in the census is required by law, and confidentiality of the information obtained is mandated by law for 72 years after collection. Census data are reported for geographic units, including states, metropolitan areas, counties, census tracts (small, relatively permanent areas within counties), and even blocks (see Exhibit 14.2). These different units allow units of analysis to be tailored to research questions. Census data are used to apportion seats in the U.S. House of Representatives and to determine federal and state legislative district boundaries, as well as to inform other decisions by government agencies.

Exhibit 14.2 Census Small-Area Geography



Source: U.S. Bureau of the Census (1994:8).

The U.S. Census website (www.census.gov) provides much information about the nearly 100 surveys and censuses that the Census Bureau directs each year, including direct access to many statistics for particular geographic units. An interactive data retrieval system, American FactFinder, is the primary means for distributing results from the 2010 Census: You can review its organization and download data at http://factfinder2.census.gov/main.html. The catalog of ICPSR (www.icpsr.umich.edu/icpsrweb/ICPSR/) also lists many census reports. Many census files containing microdata—records from persons, households, or housing units—are available online, and others can be purchased on CD-ROM or DVD from the Customer Services Center at (301) 763-INFO (4636); census data can also be inspected online or downloaded for various geographic levels, including counties, cities, census tracts, and even blocks using the DataFerrett application (Federated Electronic Research, Review, Extract, and Tabulation Tool). You can download, install, and use this tool at http://dataferrett.census.gov. This tool also provides access to data sets collected by other federal agencies. An even more accessible way to use U.S. Census data is through the website maintained by the Social Science Data Analysis Network, at www.ssdan.net. Check out the DataCounts! options.

States also maintain census bureaus and may have additional resources. Some contain the original census data collected in the state 100 or more years ago. The ambitious historical researcher can use these data to conduct detailed comparative studies at the county or state level (Lathrop 1968:79).

Integrated Public Use Microdata Series

Individual-level samples from U.S. Census data for the years 1850 to the present, as well as census files from 82 other countries from 1962 to the present, are available through the Integrated Public Use Microdata Series (IPUMS) at the University of Minnesota's Minnesota Population Center (MPC). These data are prepared in an easy-to-use format that provides consistent codes and names for all the different samples.

This exceptional resource offers samples of the U.S. population selected from 15 federal censuses, as well as results of the Census Bureau's annual American Community Survey from 2000 to 2006. Each sample is independently selected, so that individuals are not linked between samples. In addition to basic demographic measures, variables in the U.S. samples include educational, occupational, and work indicators; respondent income; disability status; immigration status; veteran status; and various household characteristics, including family composition and dwelling characteristics. Survey data are also organized in the categories of health, higher education, time use, demography and health, and the environment. The international samples include detailed characteristics from hundreds of thousands of individuals in countries ranging from France and Mexico to Kenya and Vietnam. You can view these resources at https://www.ipums.org. You must register to download data, but the registration is free.

Bureau of Labor Statistics

Another good source of data is the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor, which collects and analyzes data on employment, earnings, prices, living conditions, industrial relations, productivity and technology, and occupational safety and health (U.S. Bureau of Labor Statistics 1991, 1997b). Some of these data are collected by the U.S. Census Bureau in the monthly *CPS;* other data are collected through surveys of establishments (U.S. Bureau of Labor Statistics 1997a).

The *CPS* provides a monthly employment and unemployment record for the United States, classified by age, sex, race, and other characteristics. The *CPS* uses a stratified random sample of about 60,000 households (with separate forms for about 120,000 individuals). Detailed questions are included to determine the precise labor force status (whether they are currently working or not) of each household member over the age of 16. Statistical reports are published each month in the BLS's *Monthly Labor Review* and can also be inspected at its website (https://www.bls.gov/mlr/). Data sets are available on computer tapes and disks from the BLS and services such as ICPSR.

BLS also sponsors the National Longitudinal Surveys (NLS), which have been surveying children, youth, and adults at various ages since the mid-1960s. Measures range from education and military experience to income, mobility, health, nutrition, physical activity, fertility and sexual activity, parenting, attitudes and behaviors (https://www.bls.gov/nls/).

Other Government Sources

Many more data sets useful for historical and comparative research have been collected by federal agencies and other organizations. The National Technical Information Service (NTIS) of the U.S. Department of Commerce maintains a Federal Computer Products Center that collects and catalogs many of these data sets and related reports.

More than 3 million data sets and reports are available through the NTIS database (https://www.ntis.gov/). Data set summaries can be searched in the database by either subject or agency. Government research reports cataloged by NTIS and other agencies can be searched online at the NTIS website.

State governments are also increasingly providing data sets that researchers can download. For instance, New York's data resources may be found at https://data.ny.gov, and the state of Missouri makes census data available at http://mcdc.missouri.edu.

Other Data Sources

Many researchers who have received funding to investigate a wide range of research topics make their data available on websites where they can be downloaded by other researchers for secondary data analyses. One of the largest, introduced earlier, is the Add Health study, funded at the University of North Carolina by the National Institute of Child Health and Human Development (NICHD) and 23 other agencies and foundations to investigate influences on adolescents' health and risk behaviors (www.cpc.unc.edu/projects/addhealth). The study began in 1994–1995 with a representative sample of more than 90,000 adolescents who completed questionnaires in school and more than 20,000 who were interviewed at home. This first wave of data collection has been followed by four more, resulting in longitudinal data currently available for more than 10 years—and soon to be more than 20 years with the conclusion of the fifth wave (2016–2018). Another significant data source, the Health and Retirement Study (HRS), began in 1992 with funding from the National Institute on Aging (NIA) (http://hrsonline.isr.umich.edu). The University of Michigan oversees HRS interviews every 2 years with more than 22,000 Americans over the age of 50. To investigate family experience change, researchers at the University of Wisconsin designed the National Survey of Families and Households (http://www.ssc.wisc.edu/nsfh/). With funding from both NICHD and NIA, researchers interviewed members of more than 10,000 households in three waves, from 1987 to 2002. The Fragile Families & Child Wellbeing Study at Princeton offers a multidimensional collection of data from medical records, surveys of caregivers, parents, and children, and observation as it follows 5,000 children born in large U.S cities—many of them to unmarried parents—between 1998 and 2000 (https://fragilefamilies.princeton.edu/). The Roper Center archives public opinion poll data, with more than 23,000 U.S. and international data sets and almost 700,000 questions (and answers) from U.S. polls of adults (https://ropercenter.cornell.edu). Another noteworthy example, among many, is the Detroit Area Studies, with annual surveys between 1951 and 2004 on a wide range of personal, political, and social issues (http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/151).

Inter-university Consortium for Political and Social Research

The University of Michigan's ICPSR is the premier source of secondary data useful to social science researchers. ICPSR was founded in 1962 and now includes more than 640 colleges and universities and other institutions throughout the world. ICPSR archives the most extensive collection of social science data sets in the United States outside the federal government: More than 7,990 studies are represented in over 500,000 files from 130 countries and from sources that range from U.S. government agencies such as the Census Bureau to international organizations such as the United Nations, social research organizations such as the National Opinion Research Center, and individual social scientists who have completed funded research projects.

Types of Data Available From ICPSR

Survey data sets obtained in the United States and in many other countries that are stored at ICPSR provide data on topics ranging from elite attitudes to consumer expectations. For example, data collected in the British Social Attitudes Survey in 1998, designated by the University of Chicago's National Opinion Research Center, are available through ICPSR (go to the ICPSR website, www.icpsr.umich.edu, and search for study no. 3101). Data collected in a monthly survey of Spaniards' attitudes, by the Center for Research on Social Reality (Spain) Survey, are also available (see study no. 6964). Survey data from Russia, Germany, and other countries can also be found in the ICPSR collection.

Do you have an interest in events and interactions between nations, such as threats of military force? A data set collected by Charles McClelland includes characteristics of 91,240 such events (study no. 5211). The history of military interventions in nations around the world between 1946 and 1988 is coded in a data set developed by Frederic Pearson and Robert Baumann (study no. 6035). This data set identifies the intervener and target countries, the starting and ending dates of military intervention, and a range of potential motives (such as foreign policies, related domestic disputes, and pursuit of rebels across borders).

Census data from other nations are also available through ICPSR, as well as directly through the Internet. In the ICPSR archives, you can find a data set from the Statistical Office of the United Nations on the 1966 to 1974 population of 220 nations throughout the world (study no. 7623). More current international population data are available through data sets available from a variety of sources, such as the study of indicators of globalization from 1975 to 1995 (study no. 4172). (See also the later description of the Eurobarometer Survey Series.) More than 3,000 data sets from countries outside the United States are available through ICPSR's International Data Resource Center.

Obtaining Data From ICPSR

The data sets archived by ICPSR are available for downloading directly from the ICPSR website, www.icpsr.umich.edu. ICPSR makes data sets obtained from government sources available directly to the general public, but many other data sets are available only to individuals at the colleges and universities around the world that have paid the fees required to join ICPSR. The availability of some data sets is restricted because of confidentiality issues (see the section later in this chapter on research ethics); to use them, researchers must sign a contract and agree to certain conditions (see http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html).

You begin a search for data in the ICPSR archives at http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp. You can search the data archives for a specific topic, for specific studies (identified by study number or title), as well as for studies by specific investigators (this would be a quick way to find the data set contributed by Lawrence W. Sherman and Richard A. Berk from their research, discussed in Chapter 2, on the police response to domestic violence).

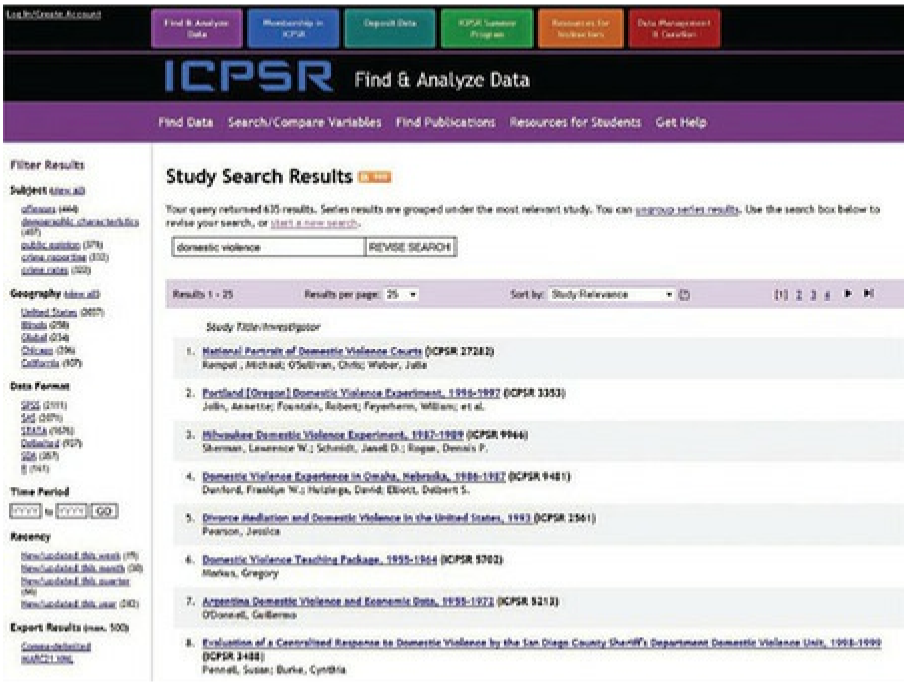
Exhibit 14.3 displays the results of a search for data sets on “domestic violence”: a list of 1,471 data sets from 444 studies that involved research on domestic violence and that are available through ICPSR. For most data sets, you can obtain a description, the files that are available for downloading, and a list of “related literature”—that is, reports and articles that use the listed data set.

When you click on the “Download” option, you are first asked to enter your e-mail address and password (ICPSR also has the option that allows sign-in with a Facebook account). What you enter will determine which data sets you can access; if you are not at an ICPSR member institution, you will be able to download only a limited portion of the data sets—mostly those from government sources. If you are a student at a member institution, you will be able to download most of the data sets directly, although you may have to be using a computer that is physically on your campus to do so.

If you prepare your own paper based on an analysis of ICPSR data, be sure to include a proper citation. Here’s an example from ICPSR itself (www.icpsr.umich.edu/icpsrweb/ICPSR/citations/):

> Reif, Karlheinz, and Anna Melich. *Euro-Barometer 39.0: European Community Policies and Family Life, March–April 1993* [Computer file]. Conducted by INRA (Europe), Brussels. ICPSR06195-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer], 1995. Koeln, Germany: Zentralarchiv fuer Empirische Sozialforschung/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 1997.

Exhibit 14.3 Search Screen: Domestic Violence Search Results



Study Search Results

Your query returned 635 results. Series results are grouped under the most relevant study. You can ungroup series results. Use the search box below to revise your search, or start a new search.

domestic violence REVISE SEARCH

Results 1 - 25 Results per page: 25 Sort by: Study Relevance

Study Title/Investigator

1. National Portrait of Domestic Violence Courts (ICPSR 27282)
Rempel, Michael; O'Sullivan, Chris; Weber, Julia
2. Portland [Oregon] Domestic Violence Experiment, 1996-1997 (ICPSR 3353)
Jolin, Annette; Fountain, Robert; Feyerherm, William; et al.
3. Milwaukee Domestic Violence Experiment, 1987-1989 (ICPSR 9966)
Sherman, Lawrence W.; Schmidt, Janell D.; Rogan, Dennis P.
4. Domestic Violence Experience in Omaha, Nebraska, 1986-1987 (ICPSR 9481)
Dunford, Franklyn W.; Huizinga, David; Elliott, Delbert S.
5. Divorce Mediation and Domestic Violence in the United States, 1993 (ICPSR 2561)
Pearson, Jessica
6. Domestic Violence Teaching Package, 1955-1964 (ICPSR 5702)
Markus, Gregory
7. Argentina Domestic Violence and Economic Data, 1955-1972 (ICPSR 5213)
O'Donnell, Guillermo
8. Evaluation of a Centralized Response to Domestic Violence by the San Diego County Sheriff's Department Domestic Violence Unit, 1998-1999 (ICPSR 3488)
Pennell, Susan; Burke, Cynthia

Source: Reprinted with permission from the Inter-University Consortium for Political and Social Research.

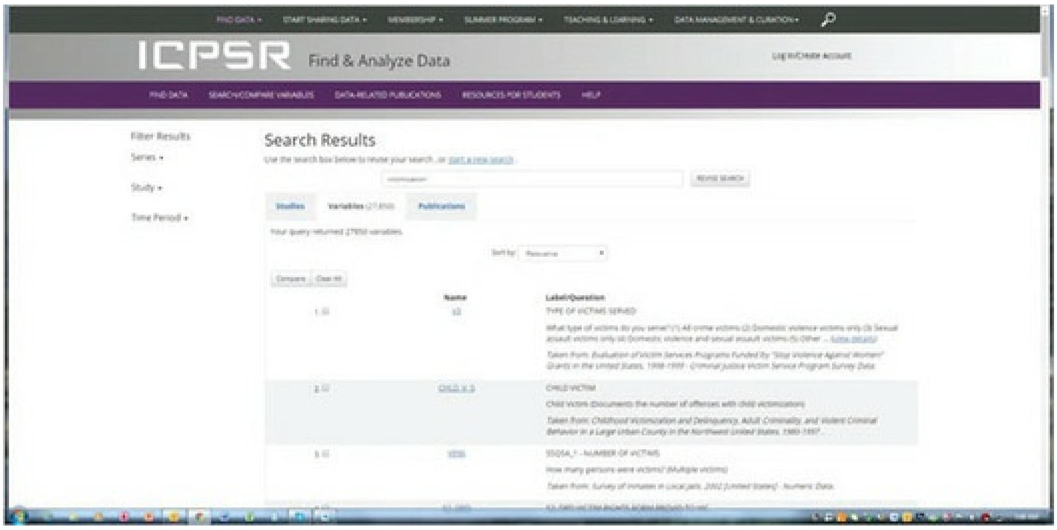
You can also search a large subset of the ICPSR database (76% of its holding) for specific variables from among the more than 4 million in this subset and identify the various studies in which they have appeared. Exhibit 14.4 displays one segment of the results of searching for variables related to "victimization." A total of 27,850 names of variables in a multitude of studies were obtained. Reviewing some of these results can suggest additional search strategies and alternative databases to consider.

Some of the data sets are also offered with the option of "Analyze Online." If you have this option, you can immediately inspect the distributions of responses to each question in a survey and examine the relation between variables, without having any special statistical programs of your own. To use this option, you must log in to ICPSR through your university library site or generate your own username and password. After doing so, you can choose the "Analyze Online" option for a data set that has this option. In Exhibit 14.5, you'll find the screen displaying the request I made with the "simple crosstab/frequency" tab for a cross-tabulation between "ever got physical with your spouse" and "police rank" in a study of police stress and domestic violence in Baltimore.

My analysis involves a cross-tabulation of the relation between likelihood of getting physical with a spouse or partner by officer rank. As you can see in Exhibit 14.6, officer rank is related to likelihood of getting physical with a spouse or partner: 1.1% of the officer trainees report they have done this, compared with 10.4% of the detectives. This approach to analysis with secondary data can jump-start your work (and may jump-start some interesting discussion as well). An online analysis option is also starting to appear at other websites that offer secondary data.

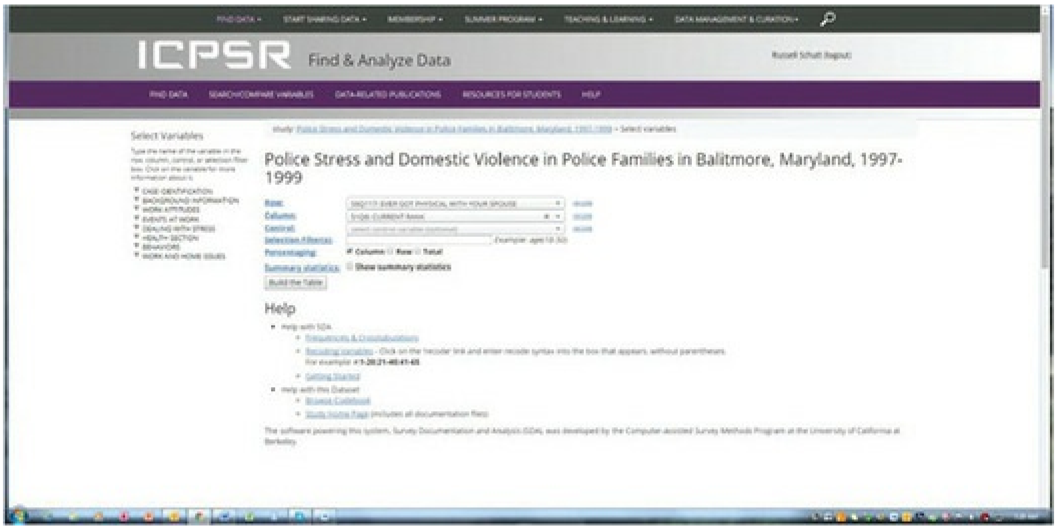
ICPSR also catalogs reports and publications containing analyses that have used ICPSR data sets since 1962—71,561 citations were in this archive on September 18, 2017. This superb resource provides an excellent starting point for the literature search that should precede a secondary data analysis. In most cases, you can learn from detailed study reports a great deal about the study methodology, including the rate of response in a sample survey and the reliability of any indexes constructed. Published articles provide examples of how others have described the study methodology as well as research questions that have already been studied with the data set and issues that remain to be resolved. You can search this literature at the ICPSR site simply by entering the same search terms that you used to find data sets or by entering the specific study number of the data set on which you have focused. Don't start a secondary analysis without reviewing such reports and publications.

Exhibit 14.4 ICPSR Variables Related to Victimization


FIND DATA
START SHARING DATA
MEMBERSHIP
SUMMER PROGRAM
TEACHING & LEARNING
DATA MANAGEMENT & CURATION
ICPSR
Find & Analyze Data
Log In/Create Account
FIND DATA
SEARCH/COMPARE VARIABLES
DATA-RELATED PUBLICATIONS
RESOURCES FOR STUDENTS
HELP
Filter Results
Series
Study
Time Period
Search Results
Studies
Variables
Publications
Name
Label/Question

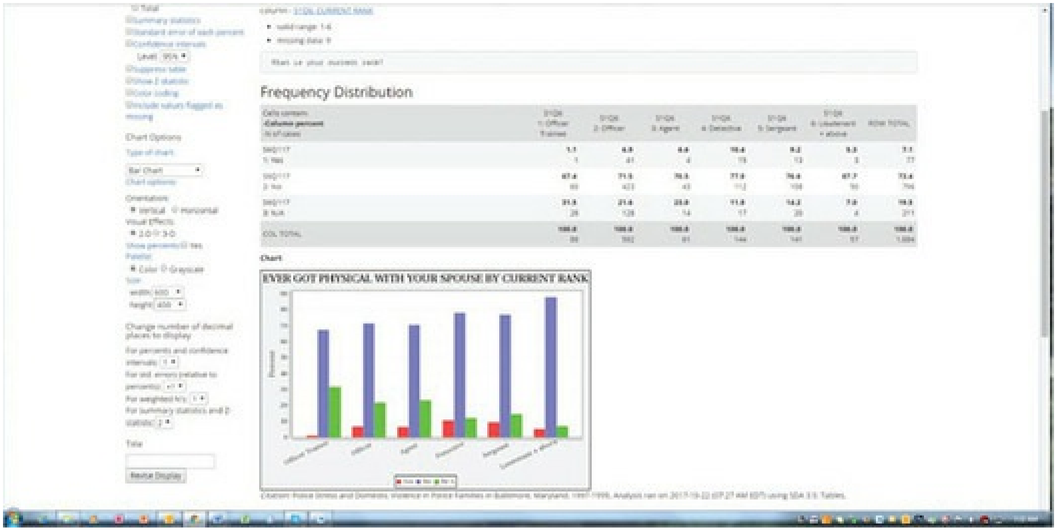
Source: Reprinted with permission from the Inter-University Consortium for Political and Social Research.

Exhibit 14.5 ICPSR Online Analysis: Codebook Information and Statistical Options


ICPSR
Find & Analyze Data
Police Stress and Domestic Violence in Police Families in Balitmore, Maryland, 1997-1999
Help

Source: Reprinted with permission from the Inter-University Consortium for Political and Social Research.

Exhibit 14.6 ICPSR Online Analysis Cross-Tabulation


Frequency Distribution
COL TOTAL
Chart
EVER GOT PHYSICAL WITH YOUR SPOUSE BY CURRENT RANK

Source: Reprinted with permission from the Inter-University Consortium for Political and Social Research.

Even if you are using ICPSR, you shouldn't stop your review of the literature with the sources listed on the ICPSR site. Conduct a search in SocINDEX or another bibliographic database to learn about related studies that used different databases (see Chapter 2).

Harvard's Dataverse

Harvard University's Henry A. Murray Research Archive (https://murray.harvard.edu) has developed a remarkable collection of social science research data sets that are now made available through a larger collaborative secondary data project as part of its Institute for Quantitative Social Science (IQSS) (https://dataverse.harvard.edu). As of September 2017, Harvard's Dataverse Project, including IQSS, provided 49,535 studies, cross-referencing many of those in the ICPSR archives. You can search data sets in the Dataverse collection by title, abstract, keywords, and other fields; if you identify a data set that you would like to analyze, you must then submit an application to be given access.

International Data Sources

Comparative researchers (using techniques we will discuss in Chapter 15) and those conducting research in other countries can find data sets on the population characteristics, economic and political features, and political events of many nations. Some of these data sets are available from U.S. government agencies. For example, the Social Security Administration reports on the characteristics of social security throughout the world (Wheeler 1995). This comprehensive source classifies nations by their type of social security program and provides detailed summaries of the characteristics of each nation's programs. Current information is available online at https://www.ssa.gov/policy/docs/progdesc/ssptw/. More recent data are organized by region. A broader range of data is available in the *World Handbook of Political and Social Indicators,* with political events and political, economic, and social data coded from 1948 to 1982 (http://www.icpsr.umich.edu, study no. 7761) (Taylor and Jodice 1986).

The European Commission administers the Eurobarometer Survey Series at least twice yearly across all the member states of the European Union. The survey monitors social and political attitudes and reports are published regularly online at https://www.gesis.org/eurobarometer-data-service/sur vey-series/. Case-level Eurobarometer survey data are stored at ICPSR. The United Nations University makes available a World Income Inequality Database from ongoing research on income inequality in developed, developing, and transition countries (http://www.wider.unu.edu/research/Database/en_GB/wiid). ICPSR also maintains an International Data Resource Center that provides access to many other data sets from around the world (https://www.icpsr.umich.edu/icpsrweb/instructors/international.jsp).

Both the Council of European Social Science Data Archives (CESSDA, https://www.cessda.eu/) and the International Federation of Data Organizations for Social Science (IFDO, ifdo.org/wordpress) maintain lists of data archives maintained by a wide range of nations (Dale, Wathan, and Higgins 2008:521). CESSDA makes available data sets from European countries to European researchers, and IFDO provides an overview of social science data sets collected throughout the world; access procedures vary, but some data sets can be downloaded directly from the IFDO site.

http://natcen.ac.uk/
http://natcen.ac.uk/our-research/research/youth-in-europe-study/
http://www.yes-deutschland.de/youth-in-europe-study/yes.html

Qualitative Data Sources

Far fewer qualitative data sets are available for secondary analysis, but the number is growing. European countries, particularly England, have been at the forefront of efforts to promote archiving of qualitative data. The United Kingdom's Economic and Social Research Council established the Qualitative Data Archiving Resource Center at the University of Essex in 1994 (Heaton 2008:507). Now part of the Economic and Social Data Service, UK Data Service QualiBank (https://www.ukdataservice.ac.uk/about-us) provides access to more than 8,000 datasets, 1,085 of which are from hundreds of qualitative or mixed-methods research projects (as of September 18, 2017). After registering at the UK Data Service site, users can browse or search directly interview transcripts and other materials from many qualitative studies, but access to many studies is restricted to users in the United Kingdom or according to other criteria.

In the United States, the ICPSR collection includes an expanding number of studies containing at least some qualitative data or measures coded from qualitative data. Studies range from transcriptions of original handwritten and published materials relating to infant and child care from the beginning of the 20th century to World War II (LaRossa 1995) to transcripts of open-ended interviews with high school students involved in violent incidents (Lockwood 1996). Harvard University's Dataverse (https://dataverse.harvard.edu) also includes qualitative studies and mixed-methods studies that contain at least some qualitative data.

The most unique source of qualitative data available for researchers in the United States is the Human Relations Area Files (HRAF) at Yale University, described in Chapter 15. The HRAF (hraf.yale.edu/resources/researchers/) has made anthropological reports available for international cross-cultural research since 1949 and contains more than 650,000 pages of information on more than 300 different cultural, ethnic, religious, and national groups (hraf.yale.edu/faq/). If you are interested in cross-cultural research, it is well worth checking out the HRAF and exploring access options (reports can be accessed and searched online by those at affiliated institutions).

The University of Southern Maine's Center for the Study of Lives (usm.maine.edu/lifestorycenter/) collects interview transcripts that record the life stories of people of diverse ages and backgrounds. Their collection includes transcripts from more than 400 life stories, representing more than 35 different ethnic groups, experiences of historical events ranging from the Great Depression to the Vietnam War, and including reports on dealing with health problems such as HIV/AIDS.

There are many other readily available sources, including administrative data from hospitals, employers, and other organizations; institutional research data from university offices that collect such data; records of transactions from businesses; and data provided directly by university-based researchers (Hakim 1982:6).

Challenges for Secondary Data Analyses

The use of the method of secondary data analysis has the following clear advantages for social researchers (Rew et al. 2000:226):

- It allows analyses of social processes in other inaccessible settings.
- It saves time and money.
- It allows the researcher to avoid data collection problems.
- It facilitates comparison with other samples.
- It may allow inclusion of many more variables and a more diverse sample than otherwise would be feasible.
- It may allow data from multiple studies to be combined.

The secondary data analyst also faces some unique challenges. The easy availability of data for secondary analysis should not obscure the fundamental differences between a secondary and a primary analysis of social science data. In fact, a researcher who can easily acquire secondary data may be tempted to minimize the limitations of the methods used to collect the data as well as insufficient correspondence between the measures in the data set and the research questions that the secondary analyst wants to answer.

So the greatest challenge faced in secondary data analysis results from the researcher's inability to design data collection methods that are best suited to answer his or her research question. The secondary data analyst also cannot test and refine the methods to be used on the basis of preliminary feedback from the population or processes to be studied. Nor is it possible for the secondary data analyst to engage in the iterative process of making observations, developing concepts, or making more observations and refining the concepts. This last problem is a special challenge for those seeking to conduct secondary analyses of qualitative data because an inductive process of developing research questions and refining observation and interview strategies is a hallmark of much qualitative methodology (Heaton 2008:511).

These limitations mean that it may not be possible for a secondary data analyst to focus on the specific research question of original interest or to use the most appropriate sampling or measurement approach for studying that research question. Secondary data analysis inevitably involves a trade-off between the ease with which the research process can be initiated and the specific hypotheses that can be tested and methods that can be used. If the primary study was not designed to measure adequately a concept that is critical to the secondary analyst's hypothesis, the study may have to be abandoned until a more adequate source of data can be found. Alternatively, hypotheses, or even the research question itself, may be modified to match the analytic possibilities presented by the available data (Riedel 2000:53).

Data quality is always a concern with secondary data, even when the data are collected by an official government agency. Government actions result, at least in part, from political processes that may not have as their first priority the design or maintenance of high-quality data for social scientific analysis. For example, political opposition to the British Census's approach to recording ethnic origin led to changes in the 1991 census that rendered its results inconsistent with prior years and that demonstrated the "tenuous relationship between enumeration [Census] categories and possible social realities" (Fenton 1996:155).

It makes sense to use official records to study the treatment of juveniles accused of illegal acts because these records document the critical decisions to arrest, to convict, or to release (Dannefer and Schutt 1982). But research based on official records can be only as good as the records themselves. In contrast to the controlled interview process in a research study, there is little guarantee that the officials' acts and decisions were recorded in a careful and unbiased manner. The same is true for data collected by employees of private and nonprofit organizations. For example, research on the quality of hospital records has created, at best, mixed support for the validity of the key information they contain (Iezzoni 1997:391).

This one example certainly does not question all legal records or all other types of official records. It does, however, highlight the value of using multiple methods, particularly when the primary method of data collection is analysis of records generated by street-level bureaucrats—officials who serve clients and have a high degree of discretion (Lipsky 1980). When officials make decisions and record the bases for their decisions without much supervision, records may diverge considerably from the decisions they are supposed to reflect. More generally, it is always important to learn how people make sense of the social world when we want to describe their circumstances and explain their behavior (Needleman 1981).

Street-level bureaucrats: Officials who serve clients and have a high degree of discretion.

The basis for concern is much greater in research across national boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Census counts can be distorted by incorrect answers to census questions as well as by inadequate coverage of the entire population (Rives and Serow 1988:32–35). National differences in the division of labor between genders within households can confuse the picture when comparing household earnings between nations without accounting for these differences (Jarvis 1997:521).

Reanalyzing qualitative data someone else collected also requires setting aside the expectation that qualitative research procedures and interpretations will be informed by intimate familiarity with the context in which the data were collected and with those from whom the data were obtained (Heaton 2008:511). Instead, the secondary analyst of qualitative data must seek opportunities for carrying on a dialogue with the original researchers.

Many of these problems can be lessened by seeking conscientiously to review data features and quality before deciding to develop an analysis of secondary data (Riedel 2000:55–69; Stewart and Kamins 1993:17–31) and then developing analysis plans that maximize the value of the available data. Replicating key analyses with alternative indicators of key concepts, testing for the stability of relationships across theoretically meaningful subsets of the data, and examining findings of comparable studies conducted with other data sets can each strengthen confidence in the findings of a secondary analysis.

Any secondary analysis will improve if the analyst—yourself or the author of the work that you are reviewing—answers several questions before deciding to develop an analysis of secondary data in the first place and then continues to develop these answers as the analysis proceeds (adapted from Riedel 2000:55–69; Stewart and Kamins 1993:17–31):

1. What were the agency's or researcher's goals in collecting the data?

   The goals of the researcher, research, or research sponsor influence every step in the process of designing a research project, analyzing the resulting data, and reporting the results. Some of these goals will be stated quite explicitly, but others may only be implicit—reflected in the decisions made but not acknowledged in the research report or other publications. When you consider whether to use a data set for a secondary analysis, you should consider whether your own research goals are similar to those of the original investigator and sponsor. The data collected are more likely to include what is necessary for achieving your own research goals if the original investigator or sponsor had similar goals. When your research question or other goals diverge from those of the original investigator, you should consider how this divergence may have affected the course of the primary research project and whether this affects your ability to use the resulting data for a different purpose.

   For example, Pamela Paxton (2002) studied the role of secondary organizations in democratic politics in a sample of 101 countries but found that she could only measure the prevalence of international nongovernmental associations (INGOs) because comparable figures on purely national associations were not available. She cautioned, "INGOs represent only a specialized subset of all the associations present in a country" (Paxton 2002:261). We need to consider this limitation when interpreting the results of her secondary analysis.
2. What data were collected, and what were they intended to measure?

   You should develop a clear description of how data enter the data collection system, for what purpose, and how cases leave the system and why. Try to obtain the guidelines that agency personnel are supposed to follow in processing cases. Have there been any changes in these procedures during the period of investigation (Riedel 2000:57–64)?

3. When was the information collected?

   Both historical and comparative analyses (see Chapter 15) can be affected. For example, the percentage of the U.S. population not counted in the U.S. Census appears to have declined since 1880 from about 7% to 1%, but undercounting continues to be more common among poorer urban dwellers and recent immigrants (King and Magnuson 1995; see also Chapter 5). The relatively successful 2000 U.S. Census reduced undercounting (Forero 2000b) but still suffered from accusations of shoddy data collection procedures in some areas (Forero 2000a).
4. What methods were used for data collection? Who was responsible for data collection, and what were their qualifications? Are they available to answer questions about the data?

   Each step in the data collection process should be charted and the involved personnel identified. The need for concern is much greater in research across national boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Incorrect answers to census questions as well as inadequate coverage of the entire population can distort census counts (see Chapter 5; Rives and Serow 1988:32–35). Copies of the forms used for data collection should be obtained, specific measures should be inspected, and the ways in which these data are processed by the agency or agencies should be reviewed.
5. How is the information organized (by date, event, etc.)? Are there identifiers that are used to identify the different types of data available (computer tapes, disks, paper files) (Riedel 2000:58–61)?

   Answers to this set of questions can have a major bearing on the work that will be needed to carry out the study.
6. What is known about the success of the data collection effort? How are missing data indicated? What kind of documentation is available? How consistent are the data with data available from other sources?

   The U.S. Census Bureau provides extensive documentation about data quality, including missing data, and it documents the efforts it makes to improve data quality. The Census 2000 Testing, Experimentation, and Evaluation Program was designed to improve the decennial census in 2010, as well as other Census Bureau censuses and surveys. This is an ongoing effort, since 1950, with tests of questionnaire design and other issues. You can read more about it at www.census.gov/pred/www/Intro.htm.

Answering these six questions helps ensure that the researcher is familiar with the data he or she will analyze and can help identify any problems with it. It is unlikely that you or any secondary data analyst will be able to develop complete answers to all these questions before starting an analysis, but it still is critical to attempt to assess what you know and don't know about data quality before deciding whether to conduct the analysis. If you uncover bases for real concern after checking documents, the other publications with the data, information on websites, and perhaps by making some phone calls, you may have to decide to reject the analytic plan and instead search for another data set. If your initial answers to these six questions give sufficient evidence that the data can reasonably be used to answer your research question, you should still keep seeking to fill in missing gaps in your initial answers to the questions; through this ongoing process, you will develop the fullest possible understanding of the quality of your data. This understanding can lead you to steer your analysis in the most productive directions and can help you write a convincing description of the data set's advantages and limitations.

This seems like a lot to ask, doesn't it? After all, you can be married for life after answering only one question; here, I'm encouraging you to attempt to answer six questions before committing yourself to a brief relationship with a data set. Fortunately, the task is not normally so daunting. If you acquire a data set for analysis from a trusted source, many of these questions will already have been answered for you. You may need to do no more than read through a description of data available on a website to answer the secondary data questions and consider yourself prepared to use the data for your own purposes. If you are going to be conducting major analyses of a data set, you should take more time to read the complete study documents, review other publications with the data, and learn about the researchers who collected the data.

Exhibit 14.7 contains the description of a data set available from ICPSR. Read through it and see how many of the secondary data questions it answers.

In an environment in which so many important social science data sets are quickly available for reanalysis, the method of secondary data analysis should permit increasingly rapid refinement of social science knowledge, as new hypotheses can be tested and methodological disputes clarified if not resolved quickly. Both the necessary technology and the supportive ideologies required for this rapid refinement have spread throughout the world. Social science researchers now have the opportunity to take advantage of this methodology as well as the responsibility to carefully and publicly delineate and acknowledge the limitations of the method.

Big Data

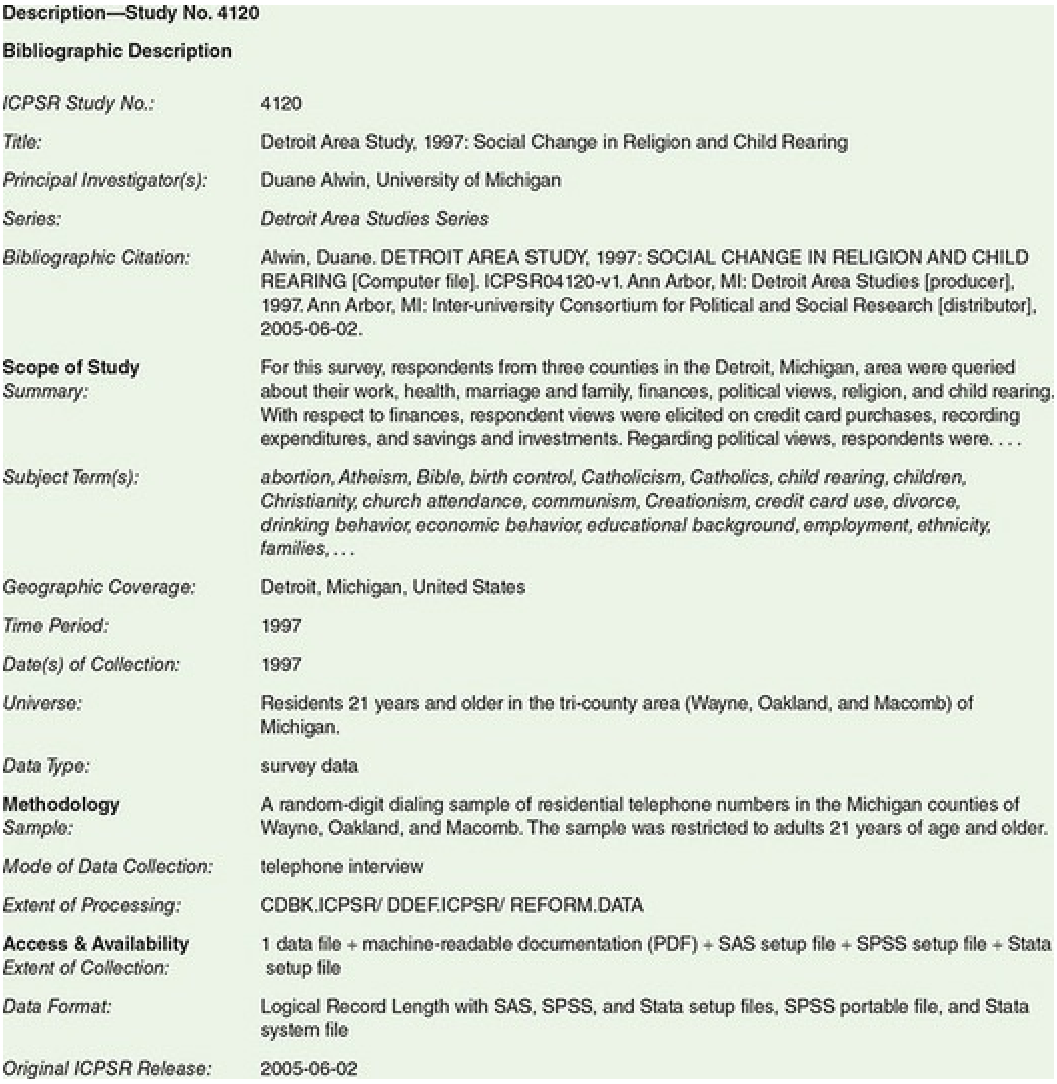
“We have come to expect information to appear when and where we need it, and to be able to communicate with anyone at anytime, no matter where they are” (Abernathy 2017:2). Do you agree? The smartphone in your hand, pocket, or purse, with its text messaging, e-mail, video chat, and other capabilities; the GPS in your car or on your phone; the social media sites you visit; the information available to you through Google searches—these all tie you in to a boundless world of information from and about people. And all of these connections leave digital tracks—“bread crumbs”—that provide data about social behavior. (Abernathy 2017:54). Big Data analyses are now being used to predict the spread of flu, the price of airline tickets, the behavior of consumers—and to investigate the social world.

Big Data: Massive data sets accessible in computer-readable form that are produced by people, available to social scientists, and manageable with today’s computers.

Background

Connections between computers began to connect people in new ways in 1969, when computers at the University of California, Los Angeles, were linked to computers at the Stanford Research Institute (and two other universities) through a network known as ARPANET—after its funder, the U.S. Defense Department's Advanced Research Projects Agency. The network, which later became known as the Internet, expanded for the next two decades with enhancements in e-mail, programming, and computer hardware, and the involvement of more organizations. In 1999, Sir Tim Berners-Lee developed the first web browser, making it much easier for users to manage their computer-mediated connections. The use of the Internet began to explode. During the same period, the U.S. Air Force created a global positioning system (GPS) by launching 24 special satellites in evenly spaced orbits that covered the globe by 1995 (now 27 are in active operation at any one time). These GPS satellites allowed two-way communication of location with units on the ground —units that now include your smartphone. The "geoweb" was born and vast amounts of interconnected data were available as never before (Abernathy 2017:19–26). And of course interpersonal interactions were changed too, as spatial media "increasingly mediate social interactions within spaces and provide different ways to know and navigate locales" (Kitchin, Lauriault, and Wilson 2017:11).

Exhibit 14.7 ICPSR Data Set Description



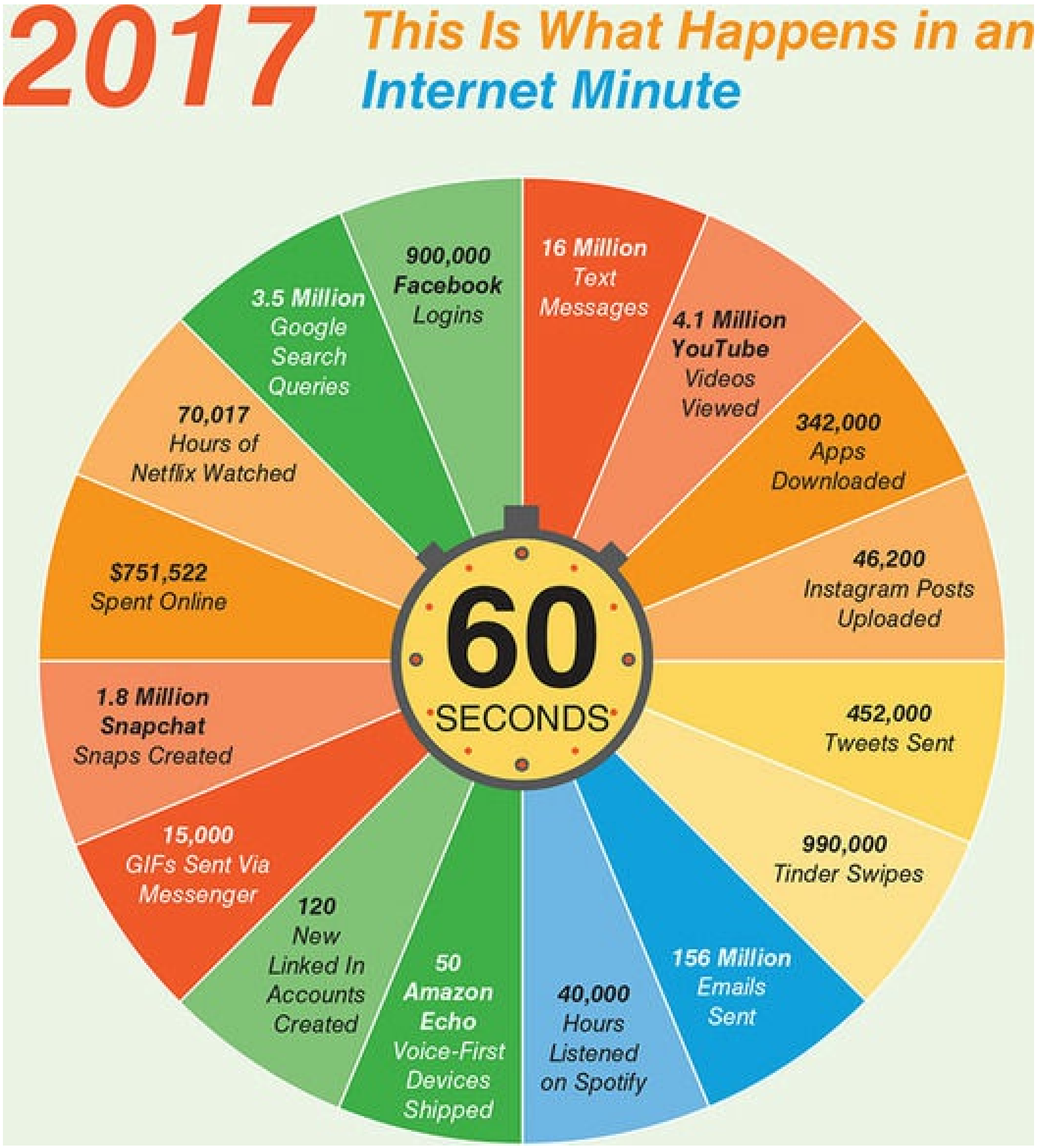
Description—Study No. 4120

Bibliographic Description

ICPSR Study No.:	4120
Title:	Detroit Area Study, 1997: Social Change in Religion and Child Rearing
Principal Investigator(s):	Duane Alwin, University of Michigan
Series:	*Detroit Area Studies Series*
Bibliographic Citation:	Alwin, Duane. DETROIT AREA STUDY, 1997: SOCIAL CHANGE IN RELIGION AND CHILD REARING [Computer file]. ICPSR04120-v1. Ann Arbor, MI: Detroit Area Studies [producer], 1997. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-06-02.
Scope of Study *Summary:*	For this survey, respondents from three counties in the Detroit, Michigan, area were queried about their work, health, marriage and family, finances, political views, religion, and child rearing. With respect to finances, respondent views were elicited on credit card purchases, recording expenditures, and savings and investments. Regarding political views, respondents were. . . .
Subject Term(s):	*abortion, Atheism, Bible, birth control, Catholicism, Catholics, child rearing, children, Christianity, church attendance, communism, Creationism, credit card use, divorce, drinking behavior, economic behavior, educational background, employment, ethnicity, families, . . .*
Geographic Coverage:	Detroit, Michigan, United States
Time Period:	1997
Date(s) of Collection:	1997
Universe:	Residents 21 years and older in the tri-county area (Wayne, Oakland, and Macomb) of Michigan.
Data Type:	survey data
Methodology *Sample:*	A random-digit dialing sample of residential telephone numbers in the Michigan counties of Wayne, Oakland, and Macomb. The sample was restricted to adults 21 years of age and older.
Mode of Data Collection:	telephone interview
Extent of Processing:	CDBK.ICPSR/ DDEF.ICPSR/ REFORM.DATA
Access & Availability *Extent of Collection:*	1 data file + machine-readable documentation (PDF) + SAS setup file + SPSS setup file + Stata setup file
Data Format:	Logical Record Length with SAS, SPSS, and Stata setup files, SPSS portable file, and Stata system file
Original ICPSR Release:	2005-06-02

Source: Detroit Area Study, 1997: Social Change in Religion and Child Rearing. Inter-University Consortium for Political and Social Research. Reprinted with permission.

Exhibit 14.8 Activity in One Internet Minute



Source: Reprinted with permission from Lori Lewis.

The quantity of data generated in this interconnected web is astounding. In September 2017, there were over 3.7 billion Internet users (over 50 percent of the world's population), more than 1.25 million websites, and almost 2 billion e-mails sent daily. Every minute, 900,000 people log into Facebook and 4.1 million videos are viewed on YouTube (for more examples, see Exhibit 14.8). Instagram users post more than 95 million photos and videos every day; and Facebook users click on a "like" button 4 million times per minute (Bagadiya 2017; Kemp 2017). Google maintains more than 1 million computer servers that process more than 60,000 search queries per second, while Twitter users send more than 500 million tweets per day (Abernathy 2017:33; http://www.internetlivestats.com/). That's "Big."

The sources of Big Data are increasing rapidly. More than two billion people use Facebook, thereby creating digital records that can, with appropriate arrangements, be analyzed to better understand social behavior (Aiden and Michel 2013:12; Desjardins 2017). Big Data are also generated by GPS users, social media, smartphones, wristband health monitors, student postings, and even student activity in online education programs (Mayer-Schönberger and Cukier 2013:90–96, 115). A new Big Data system records preemies' heart rate, respiration rate, and temperature—what amounts to 1,260 data points per second—and can predict the onset of infection 24 hours before the appearance of overt symptoms (Mayer-Schönberger and Cukier 2013:60). Public utilities, government agencies, and private companies can all learn about their customers from analyzing patterns revealed in their records. Although much of these data are inaccessible to those who are not given permission by the organizations that collect the data, and none of it is of value for those who lack the appropriate computing power and statistical skills, it is all creating possibilities for investigation of the social world that could not have been envisioned even 20 years ago.

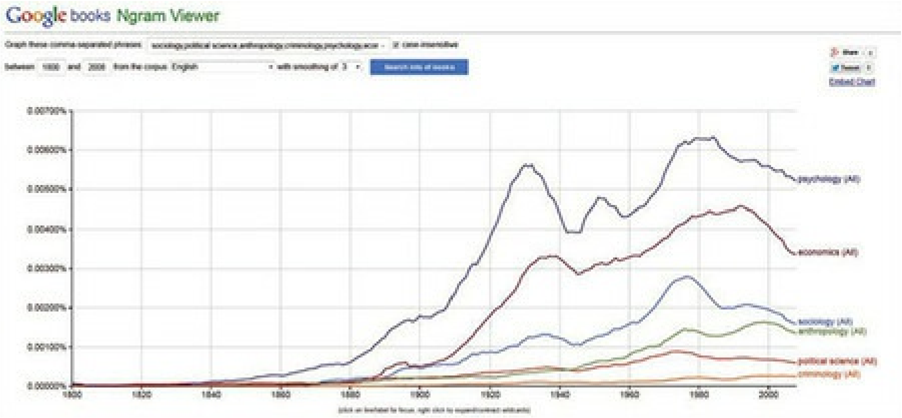
Examples of Research Using Big Data

How can you use such data in social science research? Here's a quick example. What could be more important than interest in sociology and other social sciences? After all, if you have declared sociology or another social science as your major, you probably think that this discipline is having a positive impact in the social world; maybe you want to contribute something to that impact yourself. So would you like to know how popular your discipline is? One way to answer that question is to see how frequently the name of the discipline has appeared in all the books ever written in the world. It may surprise you to learn that it is possible right now to answer that question, although with two key limitations: We can only examine books written in English and in several other languages and as of 2014 we are limited to "only" one quarter of all books ever published—a mere 30 million books (Aiden and Michel 2013:16).

To check this out, go to the Google Ngrams site (https://books.google.com/ngrams), type in "sociology, political science, anthropology, criminology, psychology, economics," and check the "case-insensitive" box (and change the ending year to 2008). Exhibit 14.9 shows the resulting screen (if you don't obtain a graph, try using a different browser). Note that the height of a graph line represents the percentage that the term represents of all words in books published in each year, so a rising line means greater relative interest in the word, not simply more books being published. You can see that psychology emerges in the mid-19th century, while sociology, economics, anthropology, and political science appear in the latter part of that century, and criminology arrives in the early 20th century. You can see that interest in sociology soared as the 1960s progressed, but then dropped off sharply in the 1980s. What else can you see in the graph? It's hard to stop checking other ideas by adding in other terms, searching in other languages, or shifting to another topic entirely. Now, that's not the same as checking how many people are reading these words in these books (*Investigating* only counts once), but it still provides quite an idea of what authors have written about. (For other limitations of Ngrams, see Zhang 2015.)

> Ngrams: Frequency graphs produced by Google's database of all words printed in more than one third of the world's books over time (with coverage still expanding).

Exhibit 14.9 Ngram of Social Sciences

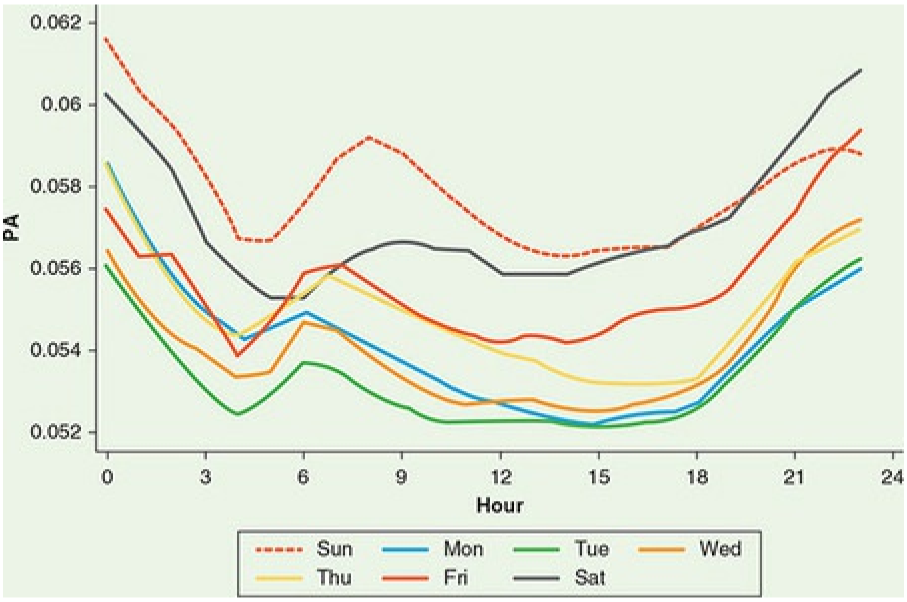


Source: Google Books Ngram Viewer, https://books.google.com/ngrams.

Dan O'Brien at Northeastern University and Chris Winship at Harvard (2017) used geographically based Big Data to investigate the importance of place in fluctuations in crime. They calculated indexes of violent crime, physical disorder, and social disorder at specific addresses from over 2 million geocoded emergency and non-emergency calls to the city of Boston's 911 (emergency) and 311 (non-emergency reports of disorder, such as pot holes or graffiti). Fewer than 1% of addresses generated 25% of reports of crime and disorder, and almost all variation in crime was between addresses. In other words, "problem properties" were responsible for most crime and disorder in Boston. You can check out the data resources and mapping capabilities that O'Brien and Winship used at the Boston Area Research Initiative (BARI) website (https://www.northeastern.edu/csshresearch/bostonarearesearchinitiative/).

Even understanding of emotions can be improved with Big Data. Sociologists Scott Golder and Michael Macy (2011) investigated mood fluctuations through the day and across the globe with 509 million messages posted on Twitter by 2.4 million individuals in 84 countries in 2008 and 2009. Using a standard system for identifying words expressing positive and negative affect (such feelings as "anxiousness," "anger," and "inhibition"), they found a common pattern of people awakening in a good mood that deteriorates throughout the day (see Exhibit 14.10). Intrigued? You too can search Tweets at https://twitter.com/search-advanced.

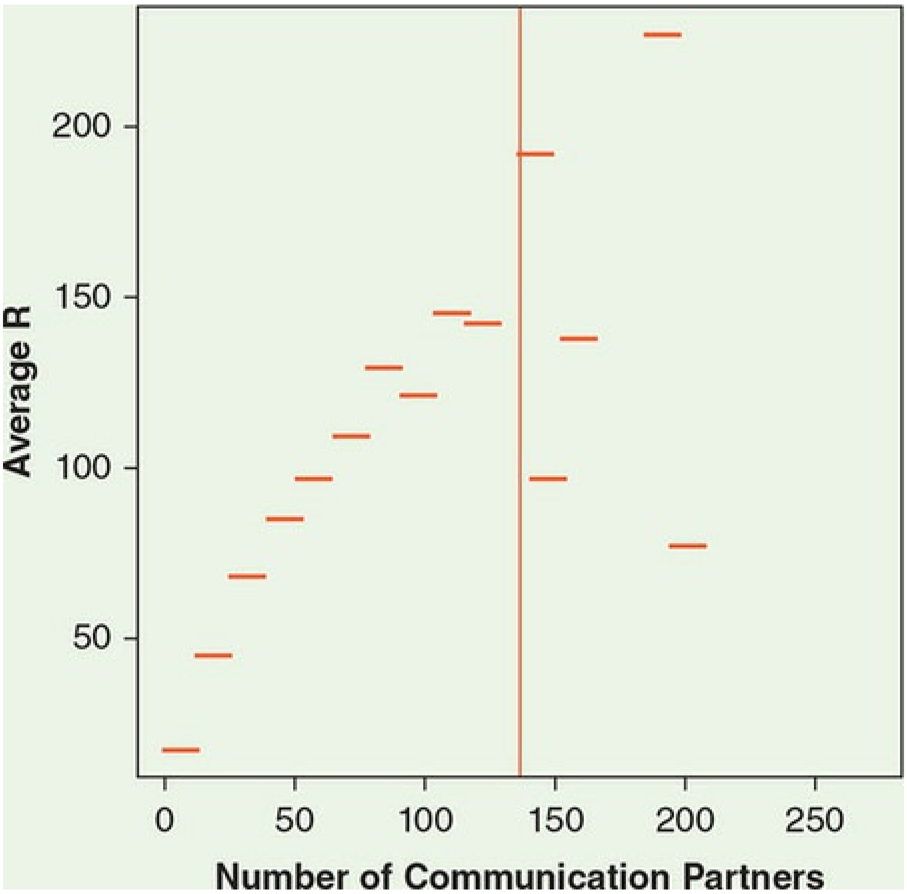
Exhibit 14.10 Hourly Changes in Individual Positive Affect (PA) by Day of the Week



Source: Golder, Scott A. and Michael W. Macy. 2011. “Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures.” Science 333:1878. Reprinted with permission from AAAS.

The availability of Big Data also makes possible the analysis of data from samples of a size previously unimaginable—even when limited research resources prevent the analysis of data from an entire population. Angela Bohn, Christian Buchta, Kurt Hornik, and Patrick Mair, in Austria and at Harvard in the United States, analyzed records on 438,851 Facebook users to explore the relation between friendship patterns and access to social capital (Bohn et al. 2014). Bohn et al. (2014:32) started their analysis with data on 1,712 users—they didn't have the computer power to analyze more data—who were selected randomly over a 2-month study period, from about 1.3 million users who had agreed on Facebook to have their data used anonymously for such a study.

Exhibit 14.11 Social Capital and Number of Communication Partners on Facebook



Source: Bohn, Angela, Christian Buchta, Kurt Hornik, and Patrick Mair. 2014. “Making friends and communicating on Facebook: Implications for the access to social capital.” Social Networks 37:29–41.

Exhibit 14.11 displays one of their findings about social networks: Having more communication partners increased social capital—as indicated by responses received to their postings—up to about 130 partners. Facebook users with more partners than that tended to receive fewer responses to their postings (Bohn et al. 2014:39). Having more partners can definitely lead to too much of what otherwise is a good thing.

In a widely discussed but cautionary tale about an attempt to use Big Data to improve public health, Jeremy Ginsberg and some colleagues at Google set out to speed up the response to the spread of flu by tracking the online searches of the 90 million U.S. adults who seek information about specific illnesses each year (Ginsberg et al. 2009:1012). By comparing the Google search data with official data collected by the U.S. Centers for Disease Control and Prevention (CDC), Ginsberg and his colleagues were able to identify search trends indicating when people first start to experience symptoms (Butler 2013:155).

But there’s also a cautionary tale here. In the 2013 flu season, Google Flu Trends predicted a much higher peak level of flu than actually occurred. The problem seems to have been that widespread media coverage and the declaration of a public health emergency in New York led many more people than usual to search for flu-related information, even though they were not experiencing symptoms themselves. Google has been refining its procedures to account for this problem, and other researchers have shifted their attention to analysis of flu-related "tweets" or to data from networks of thousands of volunteers who report symptoms experienced by family members to a central database (Butler 2013). So having incredible amounts of data does not allow us to forget about the potential for problems in measurement or sampling.

Ethical Issues in Secondary Data Analysis and Big Data

Analysis of data collected by others does not create the same potential for immediate harm as does the collection of primary data, but neither ethical nor related political considerations can be ignored. Big Data creates new ethical concerns, in part because it can reflect the behavior of many individuals who did not consent to participate in research.

In the News Research in the News: A Bright Side to Facebook's Experiments on Its Users?

For Further Thought?

Facebook, in collaboration with academic researchers, conducted a psychology experiment in 2012 with about 700,000 users. Some of these users were shown happier messages in their news feeds and others received sadder messages. This resulted in users posting their own messages with slightly happier or sadder feelings. Are you glad you now know that a site like Facebook can influence peoples' emotions? There was quite a controversy about whether Facebook had been unethical in conducting research that affected its users without their knowledge.

1. Do you believe that Facebook and other social media companies should be able to conduct academic experiments like the one described here?
2. Many businesses conduct experiments with customers or potential customers to test marketing strategies. Facebook itself has tested 41 different shades of blue on its site to see which engaged users the most. How do you compare the ethics of such research to that conducted by Facebook in order to better understand the social world? Some observers have proposed "consumer review boards" to review proposals for such research. Does this make sense to you?

News source: Manjoo, Farhad. 2014. "A Bright Side to Facebook's Experiments on Its Users." *The New York Times,* July 3, p. B1.

Subject confidentiality is a key concern when original records are analyzed with either secondary data or Big Data. Whenever possible, all information that could identify individuals should be removed from the records to be analyzed so that no link is possible to the identities of living subjects or the living descendants of subjects (Huston and Naylor 1996:1698). When you use data that have already been archived, you need to find out what procedures were used to preserve subject confidentiality. The work required to ensure subject confidentiality probably will have been done for you by the data archivist. For example, ICPSR examines carefully all data deposited in the archive for the possibility of disclosure risk. All data that might be used to identify respondents are altered to ensure confidentiality, including removal of information such as birth dates or service dates, specific incomes, or place of residence that could be used to identify subjects indirectly (see http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html). If all information that could be used in any way to identify respondents cannot be removed from a data set without diminishing data set quality (e.g., by preventing links to other essential data records), ICPSR restricts access to the data and requires that investigators agree to conditions of use that preserve subject confidentiality. Those who violate confidentiality may be subject to a scientific misconduct investigation by their home institution at the request of ICPSR (Johnson and Bullock 2009:218). The UK Data Archive provides more information about confidentiality and other human subjects protection issues at https://www.ukdataservice.ac.uk/manage-data/legal-ethical.

It is not up to you to decide whether there are any issues of concern regarding human subjects when you acquire a data set for secondary analysis from a responsible source. The institutional review board (IRB) for the protection of human subjects at your college or university or other institution has the responsibility to decide whether they need to review and approve proposals for secondary data analysis. For example, the IRB at the University of Wisconsin–Madison (UW–Madison) waives review for specific secondary data sets that meet human subjects protection standards and allows researchers to request exemption from review for other data sources that are demonstrated to meet these standards (https://kb.wisc.edu/gradsch/page.php?id=29465). Specifically, their regulations stipulate that research projects involving secondary data set analysis will *not* require prior IRB approval, if the data set has been preapproved by the UW–Madison IRB as indicated by posting on a list that includes the following data sets:

- ICPSR
- University of Wisconsin Data and Information Services Center (DISC)
- Roper Center for Public Opinion Research
- U.S. Census Bureau
- National Center for Health Statistics
- National Center for Education Statistics
- National Election Studies

Data sets that may qualify for inclusion on UW–Madison's list of approved data sources include the following:

- Public use data sets posted on the Internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects (e.g., see ICPSR's responsible use statement).
- Survey data distributed by UW principal investigators who can certify that (1) the data collection procedures were approved by a qualified IRB meeting the Common Rule criteria for IRB and (2) the data set and documentation as distributed do not contain information that could be used to identify individual research participants.

Note: Research projects that merge more than one data set in such a way that individuals may be identified are not covered by this policy and require prior IRB approval.

Data quality is always a concern with secondary data, even when the data are collected by an official government agency, and even when the data are "Big." Researchers who rely on secondary data inevitably make trade-offs between their ability to use a particular data set and the specific hypotheses they can test. If a concept that is critical to a hypothesis was not measured adequately in a secondary data source, the study might have to be abandoned until a more adequate source of data can be found. Alternatively, hypotheses or even the research question itself may be modified to match the analytic possibilities presented by the available data (Riedel 2000:53). For instance, digital data may be unrepresentative of the general population due to socioeconomic differences between those who use smartphones or connect to the Internet in other ways and those who live "offline," as well as between data sets to which we are allowed access and those that are controlled by private companies (Lewis 2015). Social behavior online may also not reflect behavior in the everyday world (Golder and Macy 2014:141–144).

Political concerns intersect with ethical practice in secondary data analyses. How are *race* and *ethnicity* coded in the U.S. Census? You learned in Chapter 4 that changing conceptualizations of race have affected what questions are asked in the census to measure race. This data collection process reflects, in part, the influence of political interest groups, and it means that analysts using the census data must understand why the proportion of individuals choosing "other" as their race and the proportion in a "multiracial" category has changed. The same types of issues influence census and other government statistics collected in other countries. Britain's census first asked about ethnic group in 1991. British researcher Steve Fenton (1996) reports that the design of the specific questions and categories used to measure ethnic groups "was clearly based on a conception of ethnic minorities as constituted by Black Caribbean and Asian populations" (p. 156). Respondents were asked to classify themselves only as white or black (in several subcategories), Indian, Pakistani, Bangladeshi, Chinese, or "any other ethnic group."

Other concerns can be much greater in research across national boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Government funding decisions can affect the availability of government statistics on particular social issues (Levitas and Guy 1996:3). Census counts can be distorted by inadequate coverage of the entire population (see Chapter 5; Rives and Serow 1988:32–35). Social and political pressures may influence the success of a census in different ways in different countries. Some Mexicans were concerned that the results of Mexico's 2000 census would be "used against them" by the government, and nearly 200,000 communities were inaccessible for follow-up except by a day's mule travel (Burke 2000). In rural China, many families that had flouted the government's official one-child policy sought to hide their "extra" children from census workers (Rosenthal 2000).

Because in most cases the secondary researchers did not collect the data, a key ethical obligation is to cite the original, principal investigators, as well as the data source, such as ICPSR. Researchers who seek access to data sets available through CESSDA must often submit a request to the national data protection authority in the country (or countries) of interest (Johnson and Bullock 2009:214).

Big Data also creates some new concerns about research ethics. When enormous amounts of data are available for analysis, the usual procedures for making data anonymous may no longer ensure that it stays that way. In 2006, AOL released for research purposes 20 million search queries from 657,000 users, after all personal information had been erased and only a unique numeric identifier remained to link searches. However, staff at *The New York Times* conducted analyses of sets of search queries and were able quickly to identify a specific individual user by name and location, based on that user's searches. The collection of Big Data also makes possible surveillance and prediction of behavior on a large scale. Crime control efforts and screening for terrorists now often involve developing predictions from patterns identified in Big Data. Without strict rules and close monitoring, potential invasions of privacy and unwarranted suspicions are enormous (Mayer-Schönberger and Cukier 2013:150–163). Should researchers be able to analyze tweets without the consent of the "tweeters" (Moe and Larsson 2012)?

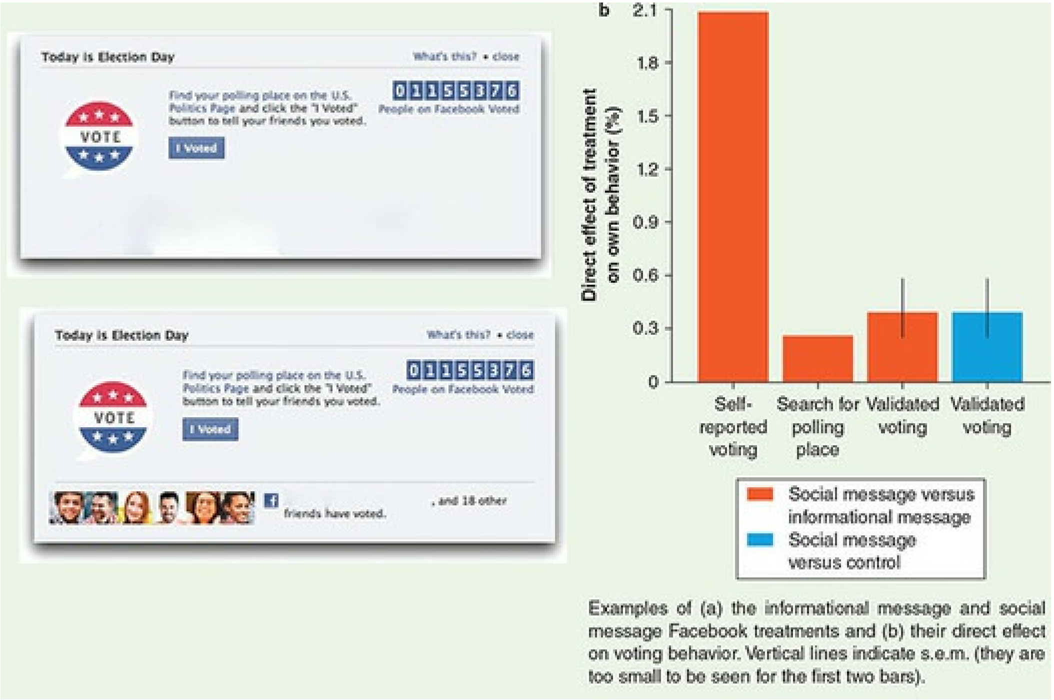
Social experiments with Big Data can literally change the social environment, and so this too raises ethical concerns. In a striking example, Robert Bond, James Fowler, and others at Facebook and the University of California, San Diego, conducted a randomized experiment with Facebook on the day of the 2010 congressional elections (Bond et al. 2012:295). Here is their description of the research design:

> Users [of Facebook] were randomly assigned to a "social message" group, an "informational message" group, or a control group. The social message group (*n* = 60,055,176) was shown a statement at the top of their "News Feed." This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading "I Voted," showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected "profile pictures" of the user's Facebook friends who had already clicked the I Voted button. . . . The informational message group (*n* = 611,044) was shown the message, poll information, counter, and button, but they were not shown any faces of friends. The control group (*n* = 613,096) did not receive any message at the top of their News Feed.

As indicated in Exhibit 14.12, individuals in the group that received the personalized message about their friends having voted were more likely to vote—and the effect was higher the more closely connected they were to those friends. Bond et al. (2012:297) estimate that receiving this message brought 60,000 more voters to the polls in 2010, whereas the effect of friends having seen the message could have increased turnout by 280,000 votes!

Do you believe that researchers should be able to conduct experiments that may alter the behavior of thousands of people? What if it changes the result of an election? Is it okay if the behavior that is increased by the experiment is widely accepted as a public good—like increasing voter turnout? How do these issues compare to those you considered in Chapter 7 about recruiting introductory psychology students for an experiment in a small group lab.

Exhibit 14.12 Facebook Experiment on Voter Turnout



Today is Election Day

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

0 1 1 5 5 3 7 6 People on Facebook Voted

Today is Election Day

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

0 1 1 5 5 3 7 6 People on Facebook Voted

, and 18 other friends have voted.

Examples of (a) the informational message and social message Facebook treatments and (b) their direct effect on voting behavior. Vertical lines indicate s.e.m. (they are too small to be seen for the first two bars).

Sources: Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D .I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." Nature 489:295–298.

Conclusions

The easy availability for secondary analyses of data sets collected in thousands of social science investigations is one of the most exciting features of social science research in the 21st century. You can often find a previously collected data set that is suitable for testing new hypotheses or exploring new issues of interest. Moreover, the research infrastructure that has developed at ICPSR and other research consortia, both in the United States and internationally, ensures that a great many of these data sets have been carefully checked for quality and archived in a form that allows easy access.

Many social scientists now review available secondary data before they consider collecting new data with which to investigate a particular research question. Even if you do not leave this course with a plan to become a social scientist yourself, you should now have the knowledge and skills required to find and use secondary data and to review analyses of Big Data to answer your own questions about the social world.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Big Data 538
Inter-university Consortium for Political and Social Research (ICPSR) 525
Ngrams 541
Secondary data 522
Secondary data analysis 522
Street-level bureaucrats 535

Highlights

- Secondary data analysts should have a good understanding of the research methods used to collect the data they analyze. Data quality is always a concern, particularly with historical data (see Chapter 15).
- ICPSR provides the most comprehensive social science data archive for academic researchers in the United States, while the U.S. Bureau of the Census and census bureaus in other countries provide official population statistics and periodic data on housing, the labor force, and many other issues.
- Collection of massive sets of Big Data permits analysis of large-scale social patterns and trends.

Discussion Questions

1. What are the strengths and weaknesses of secondary data analysis? Do you think it's best to encourage researchers to try to address their research questions with secondary data if at all possible?
2. What are the similarities and differences between secondary data analysis and Big Data analysis? Do you feel one of these approaches is more likely to yield valid conclusions? Explain your answer.
3. In a world of limited resources and time constraints, should social researchers be required to include in their proposals to collect new data an explanation of why they cannot investigate their proposed research question with secondary data? Such a requirement might include a systematic review of data that already are available at ICPSR and other sources. Discuss the merits and demerits of such a requirement. If such a requirement were to be adopted, what specific rules would you recommend?

Practice Exercises

1. Using your library's government documents collection or the U.S. Census site on the web, select one report by the U.S. Census Bureau about the population of the United States or some segment of it. Outline the report and list all the tables included in it. Summarize the report in two paragraphs.
2. Review the survey data sets available through ICPSR, using their Internet site (www.icpsr.umich.edu/icpsrweb/ICPSR/). Select two data sets that might be used to study a research question in which you are interested. Use the information ICPSR reports about them to answer the six questions in the "Challenges for Secondary Data Analyses" section of this chapter. Is the information adequate to answer these questions? What are the advantages and disadvantages of using one of these data sets to answer your research question compared with designing a new study?
3. Select a current topic and write a research question about this topic that could be answered with counts of words in books. Use the Google NGrams program described in this chapter to answer your question. Discuss the limitations of your approach, including the words you searched and the way in which you identified relationships.
4. 4. Review the "Secondary Data" lesson in the interactive exercises on the book's study site to learn more about the language and logic of secondary data analysis.

Ethics Questions

1. Reread the University of Wisconsin's IRB statement about secondary data analysis. Different policies about secondary analyses have been adopted by different IRBs. Do you agree with UW–Madison's policy? Would you recommend exempting all secondary analyses from IRB review, just some of them, or none of them? Explain your reasoning.
2. Big Data begin as little data; that is, as the records of phone calls, Twitter posts, or pictures taken by individuals in their daily lives. What limitations on access should be imposed on access to and use of such data once they have become aggregated into massive data sets? Is removing explicit identifiers sufficient protection? When does access to Big Data violate rights to privacy?
3. In January 2012, Facebook conducted an experiment in which emotional cues were manipulated for 689,003 users (see "Research in the News"). Some saw news stories and photos on Facebook's homepage containing many positive words, while others saw negative, unpleasant words. The messages sent subsequently by these users were a little more likely to reflect the emotional tone of the words they had been chosen randomly to see. When this experiment was reported in the *Proceedings of the National Academy of Sciences* (Kramer, Guillory, and Hancock 2014), some people were outraged. What do you think of the ethics of this type of Big Data experiment?

Web Exercises

1. Explore the ICPSR website. Start by browsing the list of subject headings and then write a short summary of the data sets available about one subject. You can start at www.icpsr.umich.edu/icpsrweb/ICPSR.
2. Try an online analysis. Go to one of the websites that offer online analysis of secondary data, like the ICPSR site. Review
3. the list of variables available for online analysis of one data set. Now form a hypothesis involving two of these variables, request frequency distributions for both, and generate the crosstab in percentage form to display their relationship.

Video Interview Questions

Listen to the researcher interview for Chapter 14 at edge.sagepub.com/schutt9e.

1. How is secondary data analysis being used in the motherhood and fatherhood projects?
2. Do you agree that secondary data is equally as valid as primary data? Why or why not? What are the arguments that were mentioned in the video that support the equal validity of secondary data analysis?

SPSS Exercise

This is the time to let your social scientific imagination run wild because any analysis of the GSS or ISSP data sets will qualify as a secondary data analysis. Review the list of variables in either the GSS or the ISSP, formulate one hypothesis involving two of these variables *that have no more than five categories,* and test the hypotheses using the cross-tabulation procedure. (Specify the independent variable as the column variable in the selection window, specify the dependent variable as the row variable, and then choose the Option for your table of column percents.) See if the percentage distribution of the dependent variable varies across the categories of the independent variable.

Developing a Research Proposal

If you plan a secondary data analysis research project, you will have to revisit at least one of the decisions about research designs (Exhibit 3.10, #15).

1. Convert your proposal for research using survey data, in Chapter 8, into a proposal for a secondary analysis of survey data. Begin by identifying an appropriate data set available through ICPSR. Be sure to include a statement about the limitations of your approach, in which you note any differences between what you proposed to study initially and what you are actually able to study with the available data.
2. At the ICPSR site, review the variables measured in the survey you will analyze and specify the main concepts they indicate. Specify the variables you would consider as independent and dependent in your analysis. What variables should be used as control variables?

Chapter 15 Research Using Historical and Comparative Data and Content Analysis

- Research That Matters, Questions That Count
- Overview of Historical and Comparative Research Methods
- Historical Social Science Methods
    - Historical Events Research
        - *Event-Structure Analysis*
        - *Oral History*
    - Historical Process Research
    - Cautions for Historical Methods
- Comparative Social Science Methods
- Research in the News: Britain Cracking Down on Gender Stereotypes in Ads
    - Cross-Sectional Comparative Research
- Careers and Research
    - Comparative Historical Research
    - Comparative Case Study Designs
    - Cautions for Comparative Methods
- Demographic Analysis
- Content Analysis
    - Identify a Population of Documents or Other Textual Sources
    - Determine the Units of Analysis
    - Select a Sample of Units From the Population
    - Design Coding Procedures for the Variables to Be Measured
    - Develop Appropriate Statistical Analyses
- Ethical Issues in Historical and Comparative Research and Content Analysis
- Conclusions

Although the United States and several European nations have maintained democratic systems of governance for more than 100 years, democratic rule has more often been brief and unstable, when it has occurred at all. What explains the presence of democratic practices in one country and their absence in another? Are democratic politics a realistic option for every nation? What about Libya? Egypt? Iraq? Are there some prerequisites in historical experience, cultural values, or economic resources? (Markoff 2005:384–386). A diverse set of methodological tools allows us to investigate social processes at other times and in other places, when the actual participants in these processes are not available.

Research That Matters, Questions That Count

Is an increase in democratic freedoms in nations associated with greater representation of women in powerful political positions? Prior research indicates that this is not the case; in fact, case studies have shown a drop in women's representation in government in some countries that have adopted democratic forms of governance. However, there are many complicating factors in the histories of particular nations, including whether gender quotas were implemented and the nature of the prior regime. Kathleen Fallon, Liam Swiss, and Jocelyn Viterna designed a comparative historical research project to investigate in more depth this "democracy paradox."

Fallon, Swiss, and Viterna designed a quantitative study of the "democratization process" in 118 developing countries over a 34-year period. The dependent variable in the analysis was the percentage of seats held by women in the national legislature or its equivalent. The researchers distinguished countries transitioning from civil strife, authoritarian regimes, and communist regimes, and they accounted for the use of quotas for women as well as the extent of democratic practices and the differences in national culture.

The results indicate that women's legislative representation drops after democratizing changes begin, but then increases with additional elections. However, the strength of this pattern varies with the type of pre-democratic regime and the use of quotas. The nature of the *process* of democratic change is critical to understanding its outcome for women.

1. Would you expect countries that are more democratic to have more representation of women in government? Why or why not? How would you design a research project to test your ideas?
2. Fallon, Swiss, and Viterna review separately the quantitative studies and qualitative studies conducted in the past about democratization and women's political power. What approach do you think would be most likely to improve understanding of this association? Explain your reasoning.

In this chapter, you will learn how researchers use historical and comparative methods to examine social processes over time and between countries or other large units, as well as some of the findings about democratization. By the end of the chapter, you will understand the primary steps involved in the different types of historical and comparative methods and some of the potential pitfalls of such methods. After you finish the chapter, test yourself by reading the 2012 *American Sociological Review* article by Kathleen Fallon, Liam Swiss, and Jocelyn Viterna at the *Investigating the Social World* study site and completing the related interactive exercises for Chapter 15 at edge.sagepub.com/schutt9e.

Fallon, Kathleen M., Liam Swiss, and Jocelyn Viterna. 2012. "Resolving the Democracy Paradox: Democratization and Women's Legislative Representation in Developing Nations, 1975 to 2009." *American Sociological Review* 77(3):380–408.

Historical and comparative research methods can generate new insights into social processes because of their ability to focus on aspects of the social world beyond recent events in one country. The investigation by Fallon, Swiss, and Viterna (2012) of women's representation in the legislatures of 118 developing countries over 34 years is a good example of this ability. These methods involve several different approaches and a diverse set of techniques, and they may have qualitative or quantitative components. They provide ways to investigate topics that usually cannot be studied with experiments, participant observation, or surveys. However, because this broader focus involves collecting data from records about the past or from other nations, the methods used in historical and comparative investigations present unique challenges to social researchers.

In this chapter, we will review the major methods social scientists use to understand historical processes and to compare different societies or regions. We also will introduce oral histories, a qualitative tool for historical investigations, as well as demographic methods, which can strengthen both historical and comparative studies. In addition, we will study the method of content analysis, which can be used in historical and comparative research, as well as in contemporary studies of communication with any type of media. Throughout the chapter, I will draw many examples from research on democracy and the process of democratization.

Overview of Historical and Comparative Research Methods

The central insight behind historical and comparative research is that we can improve our understanding of social processes when we make comparisons to other times and places. Max Weber's comparative study of world religions (Bendix 1962) and Émile Durkheim's (1893/2014) historical analysis of the division of labor are two examples of the central role of historical and comparative research during the period sociology emerged as a discipline. Although the popularity of this style of research ebbed with the growth of survey methods and statistical analysis in the 1930s, exemplary works such as Reinhard Bendix's (1956) *Work and Authority in Industry* and Barrington Moore Jr.'s (1966) *Social Origins of Democracy and Dictatorship* helped to fuel a resurgence of historical and comparative methods in the 1970s and 1980s that has continued into the 21st century (Lange 2013:22–33).

Historical and comparative methods are a diverse collection of approaches that can involve combinations of other methods presented in this text (Lange 2013). Research may be historical, comparative, or both historical and comparative. There are no hard-and-fast rules for determining how far in the past the focus of research must be to consider it historical or what types of comparisons are needed to warrant calling research comparative. In practice, research tends to be considered historical when it focuses on a period before the experience of most of those conducting research (Abbott 1994:80). Research involving different nations is usually considered comparative, but so are studies of different regions within one nation if they emphasize interregional comparison. In recent years, the globalization of U.S. economic ties and the internationalization of scholarship have increased the use of comparative research methods across many different countries (Kotkin 2002). Historical and comparative methods can be quantitative or qualitative, or a mixture of both. Both nomothetic and idiographic approaches to establishing causal effects can be used.

Distinguishing research with a historical or comparative focus results in four basic types of research: historical events research, historical process research, cross-sectional comparative research, and comparative historical research. Research that focuses on events in one short historical period is historical events research, whereas longitudinal research that traces a sequence of events over a number of years is historical process research (see, for example, Skocpol 1984:359). There are also two types of comparative research, the first involving cross-sectional comparisons and the second comparing longitudinal data about historical processes between multiple cases. The resulting four types of research are displayed in Exhibit 15.1.

Historical events research: Research in which social events are studied at one past time period.

Historical process research: Research in which historical processes are studied over a long time.

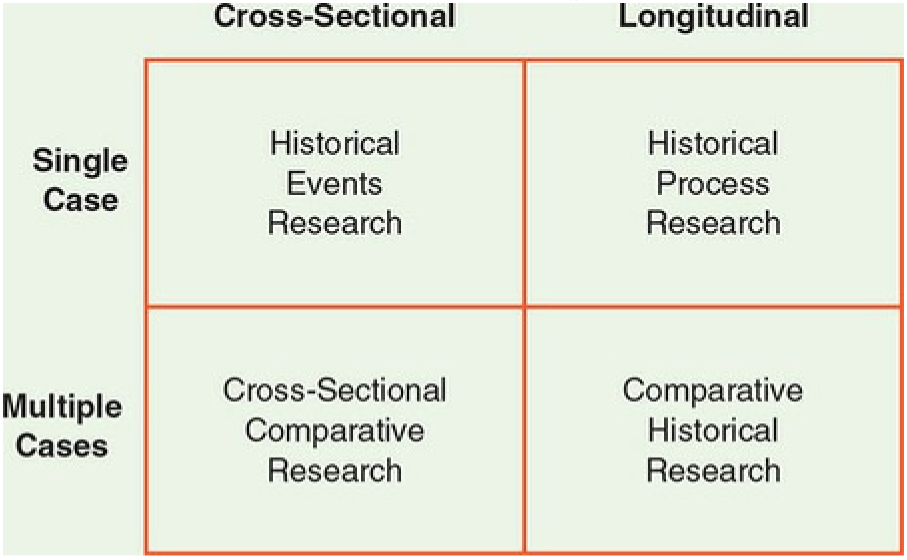
Cross-sectional comparative research: Research comparing data from one time period between two or more nations.

Comparative historical research: Research comparing data from more than one time period in more than one nation.

Historical Social Science Methods

Both historical events research and historical process research investigate questions concerning past times. These methods are used increasingly by social scientists in sociology, anthropology, political science, and economics, as well as by many historians (Monkkonen 1994). The late 20th and early 21st centuries have seen so much change in so many countries that many scholars have felt a need to investigate the background of these changes and to refine their methods of investigation (Hallinan 1997; Robertson 1993). The accumulation of large bodies of data about the past not only has stimulated more historically oriented research but also has led to the development of several different methodologies.

Exhibit 15.1 Types of Historical and Comparative Research



	Cross-Sectional	**Longitudinal**
Single Case	Historical Events Research	Historical Process Research
Multiple Cases	Cross-Sectional Comparative Research	Comparative Historical Research

Much historical (and comparative) research is qualitative. This style of historical social science research tends to have several features that are similar to those used in other qualitative methodologies. First, like other qualitative methods, qualitative historical research is inductive: it develops an explanation for what happened from the details discovered about the past. In addition, qualitative historical research is case oriented; it focuses on the nation or other unit as a whole, rather than only on different parts of the whole in isolation from each other (Ragin 2000:68). This could be considered the most distinctive feature of qualitative research on historical processes. The research question is "What was Britain like at the time?" rather than "What did Queen Elizabeth do?" Related to this case orientation, qualitative historical research is holistic—concerned with the context in which events occurred and the interrelations between different events and processes: "how different conditions or parts fit together" (Ragin 1987:25–26). For the same reason, qualitative historical research is conjunctural because, it is argued, “no cause ever acts except in complex conjunctions with others” (Abbott 1994:101). Charles Ragin (2000:67–68) uses the example of case-oriented research on the changing relationship between income and single parenthood in the United States after World War II:

> In the end, the study is also about the United States in the second half of the twentieth century, not just the many individuals and families included in the analysis. More than likely, the explanation of the changing relation between income and single parenthood would focus on interrelated aspects of the United States over this period, For example, to explain the weakening link between low income and single parenthood the researcher might cite the changing status of women, the decline in the social significance of conventional family forms, the increase in divorce, the decrease in men’s job security, and other changes occurring in the United States over this period.

Qualitative historical research is also temporal because it looks at the related series of events that unfold over time. It is therefore also likely to be *historically specific*—limited to the specific time(s) and place(s) studied. Qualitative historical research uses narrative explanations—idiographic causal reasoning (see Chapter 6)—in which the research tells a story involving specific actors and other events occurring at the same time (Abbott 1994:102) or one that accounts for the position of actors and events in time and in a unique historical context (Griffin 1992). Larry Griffin’s (1993) research on lynching, in the next section, provides a good example.

Case-oriented research: Research that focuses attention on the nation or other unit as a whole.

Holistic research: Research concerned with the context in which events occurred and the interrelations between different events and processes.

Conjunctural research: Research that considers the complex combinations in which causal influences operate.

Temporal research: Research that accounts for the related series of events that unfold over time.

Narrative explanation: An idiographic causal explanation that involves developing a narrative of events and processes that indicate a chain of causes and effects.

The focus on the past presents special methodological challenges:

- Documents and other evidence may have been lost or damaged.
- Available evidence may represent a sample biased toward more newsworthy figures.
- Written records will be biased toward those who were more prone to writing.

- Feelings of individuals involved in past events may be hard, if not impossible, to reconstruct.

Before you judge historical social science research as credible, you should look for convincing evidence that each of these challenges has been addressed.

Historical Events Research

Research on past events that does not follow processes for some long period is historical events research rather than historical process research. Historical events research basically uses a cross-sectional, rather than longitudinal, design. Investigations of past events may be motivated by the belief that they had a critical impact on subsequent developments or because they provide opportunities for testing the implications of a general theory (Kohn 1987).

Event-Structure Analysis

One technique useful in historical events research, as well as in other types of historical and comparative research, is event-structure analysis. Event-structure analysis is a qualitative approach that relies on a systematic coding of key events or national characteristics to identify the underlying structure of action in a chronology of events. The codes are then used to construct event sequences, make comparisons between cases, and develop an idiographic causal explanation for a key event.

An event-structure analysis consists of the following steps:

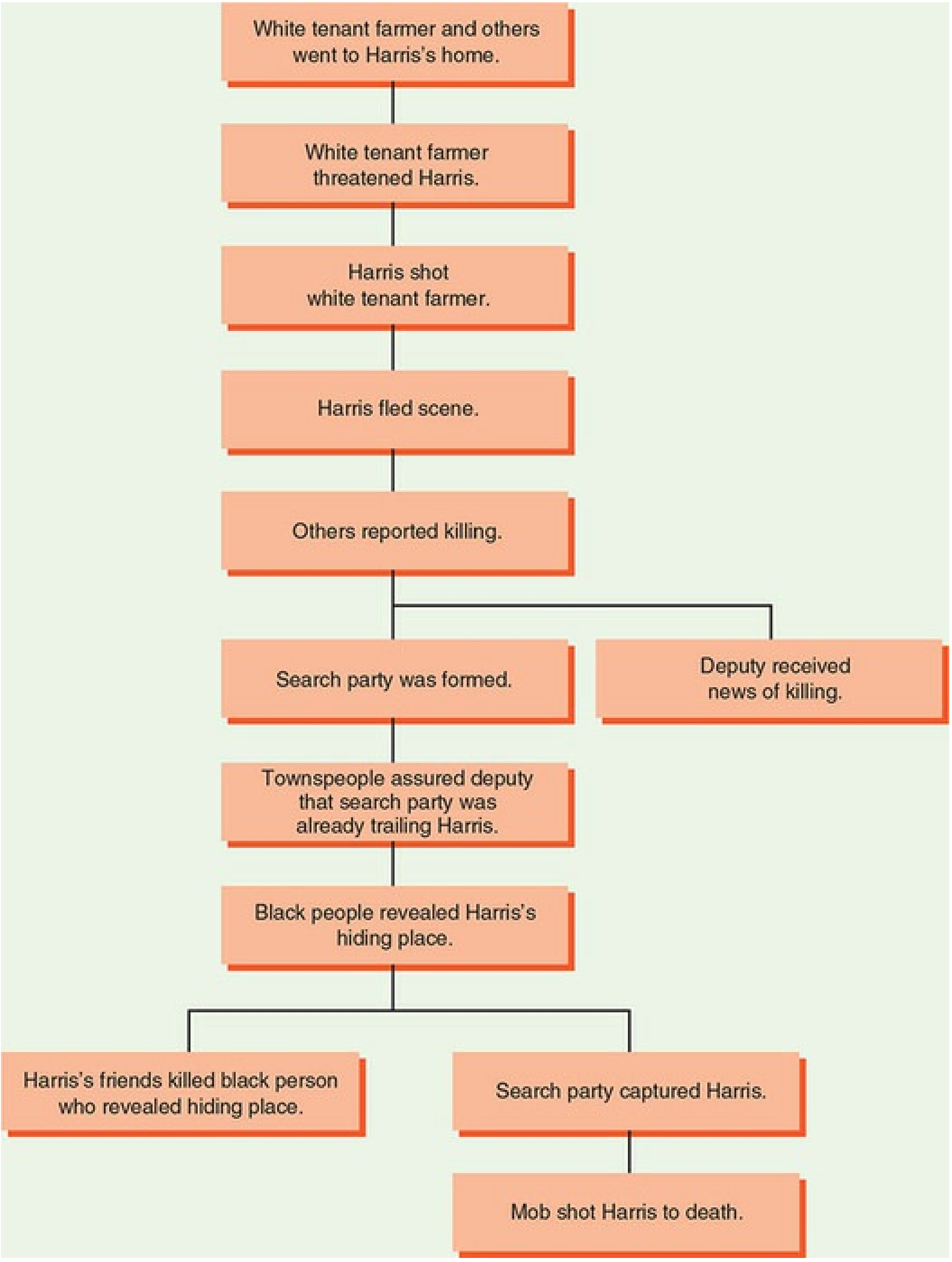
1. Classifying historical information into discrete events
2. Ordering events into a temporal sequence
3. Identifying prior steps that are prerequisites for subsequent events
4. Representing connections between events in a diagram
5. Eliminating from the diagram connections that are not necessary to explain the focal event

Griffin (1993) used event-structure analysis to explain a unique historical event, a lynching in the 1930s in Mississippi. According to published accounts and legal records, the lynching occurred after David Harris, an African American who sold moonshine from his home, was accused of killing a white tenant farmer. After the killing was reported, the local deputy was called and a citizen search party was formed. The deputy did not intervene as the search party trailed Harris and then captured and killed him. Meanwhile, Harris's friends killed another African American who had revealed Harris's hiding place. This series of events is outlined in Exhibit 15.2.

Event-structure analysis: A systematic method of developing a causal diagram showing the structure of action underlying some chronology of events; the result is an idiographic causal explanation.

Which among the numerous events occurring between the time that the tenant farmer confronted Harris and the time that the mob killed Harris had a causal influence on that outcome? To identify these idiographic causal links, Griffin identified plausible counterfactual possibilities—events that might have occurred but did not—and considered whether the outcome might have been changed if a counterfactual had occurred instead of a particular event (see Chapter 6):

Exhibit 15.2 Event-Structure Analysis: Lynching Incident in the 1930s


White tenant farmer and others went to Harris's home.
White tenant farmer threatened Harris.
Harris shot white tenant farmer.
Harris fled scene.
Others reported killing.
Search party was formed.
Deputy received news of killing.
Townspeople assured deputy that search party was already trailing Harris.
Black people revealed Harris's hiding place.
Harris's friends killed black person who revealed hiding place.
Search party captured Harris.
Mob shot Harris to death.

Source: Adapted from Griffin (1993:1110).

> If, contrary to what actually happened, the deputy had attempted to stop the mob, might the lynching have been averted? . . . Given what happened in comparable cases and the Bolivar County deputy's clear knowledge of the existence of the mob and of its early activities, his forceful intervention to prevent the lynching thus appears an objective possibility. (Griffin 1993:1112)

So Griffin concluded that nonintervention by the deputy had a causal influence on the lynching.

Oral History

History that is not written down is mostly lost to posterity (and social researchers). However, oral histories can be useful for understanding historical events that occurred within the lifetimes of living individuals. As the next example shows, sometimes oral histories even result in a written record that can be analyzed by researchers at a later point in time.

Thanks to a Depression-era writers' project, Deanna Pagnini and Philip Morgan (1996) found that they could use oral histories to study attitudes toward out-of-wedlock births among African American and white women in the South during the 1930s.

Almost 70% of African American babies are born to unmarried mothers, compared with 22% of white babies (Pagnini and Morgan 1996:1696). This difference often is attributed to contemporary welfare policies or problems in the inner city, but Pagnini and Morgan thought it might be due to more enduring racial differences in marriage and childbearing. To investigate these historical differences, they read 1,170 life histories recorded by almost 200 writers who worked for a New Deal program during the Depression of the 1930s, the Federal Writers' Project southern life history program. The interviewers had used a topic outline that included family issues, education, income, occupation, religion, medical needs, and diet.

In 1936, the divergence in rates of nonmarital births was substantial in North Carolina: 2.6% of white births were to unmarried women, compared with 28.3% of nonwhite births. The oral histories gave some qualitative insight into community norms that were associated with these patterns. A white seamstress who became pregnant at age 16 recalled, "I'm afraid he didn't want much to marry me, but my mother's threats brought him around" (Pagnini and Morgan 1996:1705). There were some reports of suicides by unwed young white women who were pregnant. In comparison, African American women who became pregnant before they were married reported regrets, but rarely shame or disgrace. There were no instances of young black women dying by suicide or getting abortions in these circumstances.

> We found that bearing a child outside a marital relationship was clearly not the stigmatizing event for African Americans that it was for whites. . . . When we examine contemporary family patterns, it is important to remember that neither current marriage nor current childbearing patterns are “new” for either race. Our explanations for why African Americans and whites organize their families in different manners must take into account past behaviors and values. (Pagnini and Morgan 1996:1714–1715)

Whether oral histories are collected by the researcher or obtained from an earlier project, the stories they tell can be no more reliable than the memories that are recalled. Unfortunately, memories of past attitudes are “notoriously subject to modifications over time” (Banks 1972:67), as are memories about past events, relationships, and actions. Use of corroborating data from documents or other sources should be used when possible to increase the credibility of descriptions based on oral histories.

Oral history: Data collected through intensive interviews with participants in past events.

Historical Process Research

Historical process research extends historical events research by focusing on a series of events that happened over a longer period. This longitudinal component allows for a much more complete understanding of historical developments than is often the case with historical events research, although it often uses techniques that are also used for research on historical events at one point in time, such as event history analysis and oral histories.

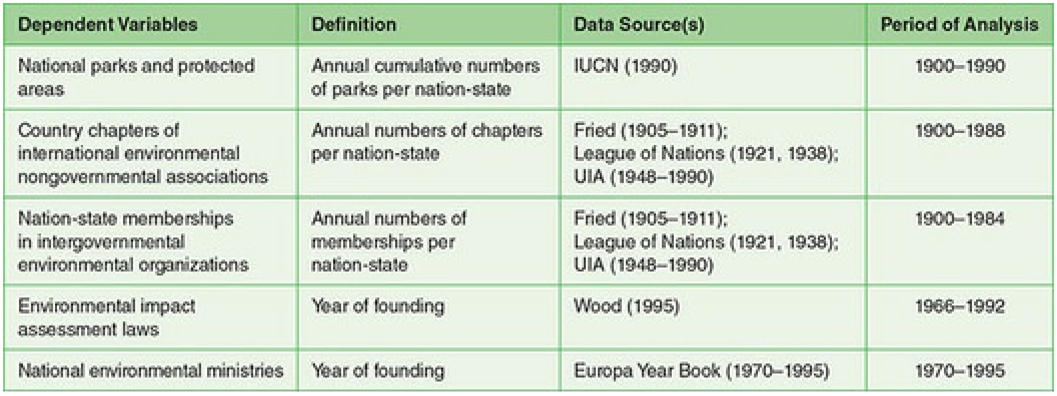
Historical process research can also use quantitative techniques. The units of analysis in quantitative analyses of historical processes are nations or larger entities, and researchers use a longitudinal design to identify changes over time. For example, David John Frank, Ann Hironaka, and Evan Schofer (2000) treated the entire world as their "case" for their deductive test of alternative explanations for the growth of national activities to protect the natural environment during the 20th century. Were environmental protection activities a response to environmental degradation and economic affluence within nations, as many had theorized? Or, instead, were they the result of a "top-down" process in which a new view of national responsibilities was spread by international organizations? Frank et al.'s measures of environmental protectionism included the number of national parks among all countries in the world and memberships in international environmental organizations; one of their indicators of global changes was the cumulative number of international agreements (see Exhibit 15.3 for a list of some of their data sources).

Exhibit 15.4a charts the growth of environmental activities identified around the world. Compare the pattern in this exhibit with the pattern of growth in the number of international environmental agreements and national environmental laws shown in Exhibit 15.4b, and you can see that environmental protectionism at the national level was rising at the same time that it was becoming more the norm in international relations. In more detailed analyses, Frank and colleagues (2000) attempt to show that the growth in environmental protectionism was not explained by increasing environmental problems or economic affluence within nations. As in most research that relies on historical or comparative data, however, some variables that would indicate alternative influences (such as the strength of national environmental protest movements) could not be measured (Buttel 2000). Therefore, further research is needed.

Cautions for Historical Methods

One common measurement problem in historical research projects is the lack of data from some historical periods (Rueschemeyer, Stephens, and Stephens 1992:4; Walters, James, and McCammon 1997). For example, the widely used U.S. Uniform Crime Reporting System did not begin until 1930 (Rosen 1995). Sometimes, alternative sources of documents or estimates for missing quantitative data can fill in gaps (Zaret 1996), but even when measures can be created for key concepts, multiple measures of the same concepts are likely to be out of the question; as a result, tests of reliability and validity may not be feasible. Whatever the situation, researchers must assess the problem honestly and openly (Bollen, Entwisle, and Alderson 1993; Paxton 2002).

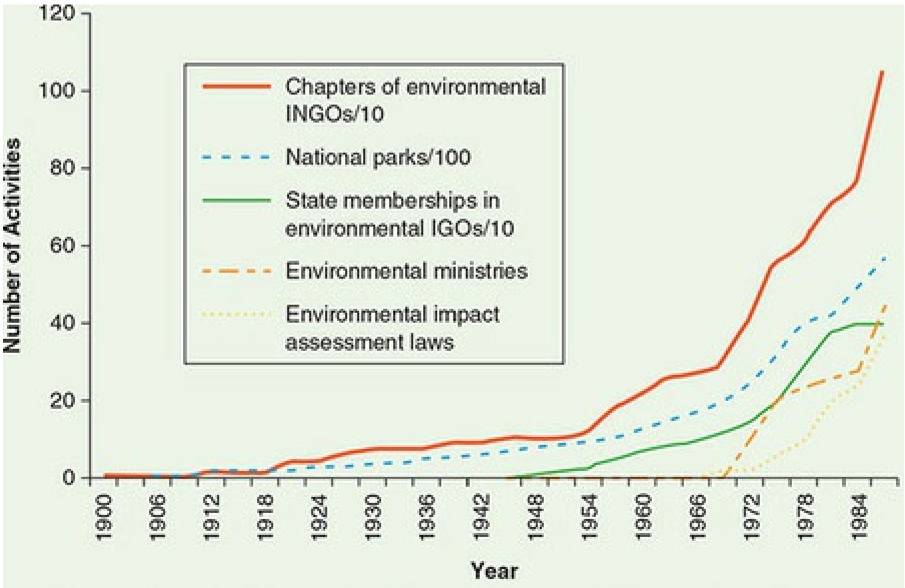
Exhibit 15.3 Variables for Historical Analysis of Environmental Protectionism



Dependent Variables	Definition	Data Source(s)	Period of Analysis
National parks and protected areas	Annual cumulative numbers of parks per nation-state	IUCN (1990)	1900–1990
Country chapters of international environmental nongovernmental associations	Annual numbers of chapters per nation-state	Fried (1905–1911); League of Nations (1921, 1938); UIA (1948–1990)	1900–1988
Nation-state memberships in intergovernmental environmental organizations	Annual numbers of memberships per nation-state	Fried (1905–1911); League of Nations (1921, 1938); UIA (1948–1990)	1900–1984
Environmental impact assessment laws	Year of founding	Wood (1995)	1966–1992
National environmental ministries	Year of founding	Europa Year Book (1970–1995)	1970–1995

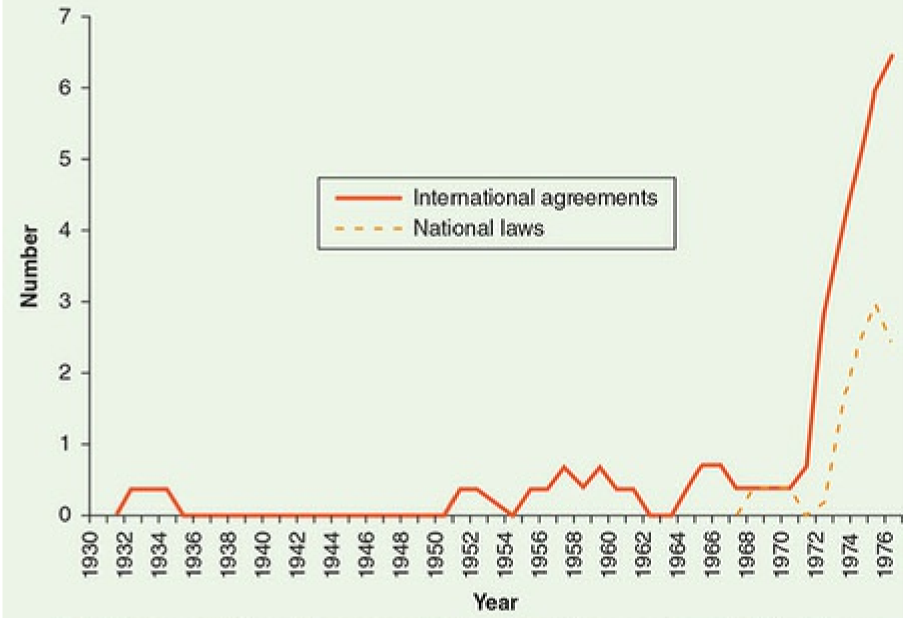
Source: “The Nation-State and the Natural Environment over the Twentieth Century.” David John Frank, Ann Hironaka, and Evan Schofer. February 2000: Vol. 65, No.1, pp. 96–116. American Sociological Review.

Exhibit 15.4 International Environmental Activity



(a) Cumulative Numbers of Five National Environment Activities, 1900–1988

Note: INGOs are international nongovernment organizations; IGOs are intergovernmental organizations.



(b) Environmental Impact Assessment: International Agreements and National Laws, 1930–1977

Source: "The Nation-State and the Natural Environment over the Twentieth Century." David John Frank, Ann Hironaka, and Evan Schofer. February 2000: Vol. 65, No.1, pp. 96–116. American Sociological Review.

Those measures that are available are not always adequate. What is included in the historical archives may be an unrepresentative selection of materials that remain from the past. At various times, some documents could have been discarded, lost, or transferred elsewhere for a variety of reasons. *Original* documents may be transcriptions of spoken words or handwritten pages and could have been modified slightly in the process; they could also be outright distortions (Erikson 1966:172, 209–210; Zaret 1996). When relevant data are obtained from previous publications, it is easy to overlook problems of data quality, but this simply makes it all the more important to evaluate the primary sources. It is very important to develop a systematic plan for identifying and evaluating relevant documents.

Comparative Social Science Methods

The limitations of single-case historical research have encouraged many social scientists to turn to comparisons between nations. These studies allow for a broader vision about social relations than is possible with cross-sectional research limited to one country or another unit. From 1985 to 1990, more than 80 research articles in top sociology journals and 200 non-edited books were published in which the primary purpose was the comparison of two or more nations (Bollen et al. 1993). About half of this research used cross-sectional data rather than longitudinal data collected over time. Both cross-sectional and longitudinal comparative methods encourage testing the cross-population generalizability of findings based on only one context (see Chapter 2).

In the News Research in the News: Britain Cracking Down on Gender Stereotypes in Ads

For Further Thought?

Britain’s Advertising Standards Authority reported that gender stereotypes in ads could “restrict the choices, aspirations and opportunities” of girls and teenagers and others who view the ads. It is developing new standards for advertising that it will then enforce. Ads that fail “to demonstrate the mother’s value to the family” or otherwise endorse gender equality could be banned. Feminist groups, marketing groups, and journalists are debating the proposed standards.

1. What are the expectations about gender equality in your country? Can you imagine rules like those under consideration in Britain being endorsed there?
2. What indicators of gender inequality would you propose for historical and comparative research?

News source: Magra, Iliana. 2017. “Britain Cracking Down on Gender Stereotypes in Ads.” *The New York Times,* July 18.

Cross-Sectional Comparative Research

Comparisons between countries during one time period can help social scientists identify the limitations of explanations based on single-nation research. Such comparisons can suggest the relative importance of universal factors in explaining social phenomena compared with unique factors rooted in specific times and places (de Vaus 2008:251). These comparative studies may focus on a period in either the past or the present. Peter Houtzager and Arnab Acharya (2011) also point out that it can be more appropriate to compare cities or regions when the nations in which they are embedded vary internally in their social characteristics. For example, they compare the impact of engagement in associations on citizenship activity in São Paolo, Brazil, and Mexico City because the conditions exist for such an impact in these cities, rather than in the surrounding countries.

Historical and comparative research that is quantitative may obtain data from national statistics or other sources of published data; if it is contemporary, such research may rely on cross-national surveys. Like other types of quantitative research, quantitative historical and comparative research can be termed variable-oriented research, with a focus on variables representing particular aspects of the units studied (Demos 1998).

Causal reasoning in quantitative historical and comparative research is nomothetic, and the approach is usually deductive, testing explicit hypotheses about relations between these variables (Kiser and Hechter 1991). For example, Clem Brooks and Jeff Manza (2006:476–479) deduce from three theories about welfare states—national values, power resources, and path dependency theory—the hypothesis that voters' social policy preferences will influence welfare state expenditures. Using country-level survey data collected by the International Social Survey Program (ISSP) in 15 democracies in five different years and expenditure data from the Organisation for Economic Co-operation and Development (OECD), Brooks and Manza were able to identify a consistent relationship between popular preferences for social welfare spending and the actual national expenditures (see Exhibit 15.5).

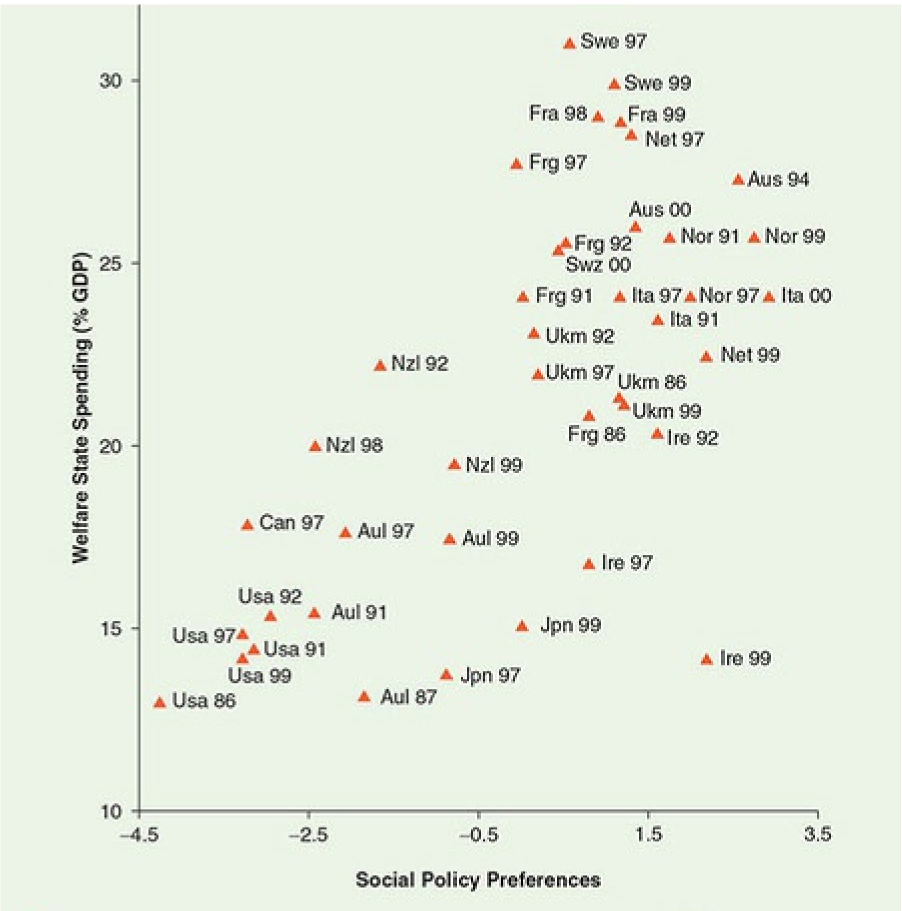
Popular preferences are also important factors in political debates over immigration policy. Christopher Bail (2008) asked whether majority groups in different European countries differ in the way that they construct "symbolic boundaries" that define "us" versus an immigrant "them." For his cross-sectional comparative investigation, he drew on 333,258 respondents in the 21-country European Social Survey (ESS). The key question about immigrants in the ESS was "Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside [country] should be able to come and live here." The "things" whose importance they were asked to rate were six individual characteristics: (1) being white, (2) being well educated, (3) being from a Christian background, (4) speaking the official national language, (5) being committed to the country's way of life, and (6) having work skills needed in the country.

Bail then calculated the average importance rating in each country for each of these characteristics and used a statistical procedure to cluster the countries by the extent to which their ratings and other characteristics were similar.

Bail's (2008:54–56) analysis identified the countries as falling into three clusters (see Exhibit 15.6). Cluster A countries are on the periphery of Europe and have only recently experienced considerable immigration; their populations tend to draw boundaries by race and religion. Cluster B countries are in the core of Western Europe (except Slovenia), they have a sizable and long-standing immigrant population, and their populations tend to base their orientations toward immigrants on linguistic and cultural differences. Countries in Cluster C are in Scandinavia, have a varied but relatively large immigrant population, and attach much less importance to any of the six symbolic boundaries than do those in the other countries. Bail (2008:56) encourages longitudinal research to determine the extent to which these different symbolic boundaries are the product or the source of social inequality in these countries.

Variable-oriented research: Research that focuses attention on variables representing particular aspects of the cases studied and then examines the relations between these variables across sets of cases.

Exhibit 15.5 Interrelationship of Policy Preferences and Welfare State Output

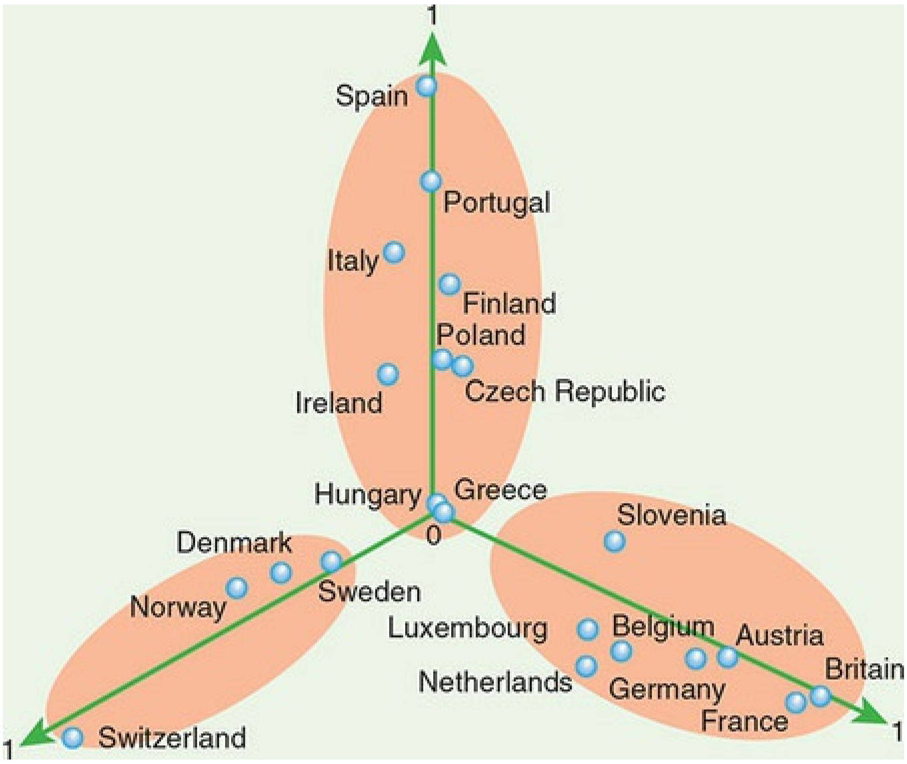


Note: Scattergram shows data for policy preferences and welfare state spending in 15 OECD democracies. Data are from the International Social Survey Program/Organisation for Economic Co-operation and Development.

Source: Brooks, Clem and Jeff Manza. June, 2006. “Social Policy Responsiveness in Developed Democracies.” American Sociological Review 71(3):474–494.

Cross-sectional comparative research has also helped explain variation in voter turnout. This research focuses on a critical issue in political science: Although free and competitive elections are a defining feature of democratic politics, elections themselves cannot orient governments to popular sentiment if citizens do not vote (LeDuc, Niemi, and Norris 1996). As a result, the low levels of voter participation in U.S. elections have long been a source of practical concern and research interest. International data give our first clue for explaining voter turnout: The historic rate of voter participation in the United States (48.3%, on average) is much lower than it is in many other countries that have free, competitive elections; for example, Italy has had a voter turnout of 92.5%, on average, since 1945 (see Exhibit 15.7).

Exhibit 15.6 Symbolic Boundaries Against Immigrants in 21 European Countries



Source: Bail, Christopher A. February 2008. “The Configuration of Symbolic Boundaries against Immigrants in Europe.” American Sociological Review 73(1):37–59.

Does this variation result from differences between voters in knowledge and wealth? Do media and political party get-out-the-vote efforts matter? Mark Franklin’s (1996:219–222) analysis of international voting data indicates that neither explanation accounts for much of the international variation in voter turnout. Instead, the structure of competition and the importance of issues are influential. Voter turnout is maximized where structural features maximize competition: compulsory voting (including, in Exhibit 15.7, Austria, Belgium, Australia, and Greece), mail and Sunday voting (including the Netherlands and Germany), and multiday voting. Voter turnout also tends to be higher where the issues being voted on are important and where results are decided by proportional representation (as in Italy and Israel, in Exhibit 15.7) rather than on a winner-take-all basis (as in U.S. presidential elections)—so individual votes are more important.

Franklin concludes that it is these characteristics that explain the low level of voter turnout in the United States, rather than the characteristics of individual voters. The United States lacks the structural features that make voting easier, the proportional representation that increases the impact of individuals’ votes, and, often, the sharp differences between candidates that are found in countries with higher turnout. Because these structural factors generally do not vary within nations, we would never realize their importance if our analysis was limited to data from individuals in one nation.

Despite the unique value of comparative analyses like Franklin's (1996), such cross-national research also confronts unique challenges (de Vaus 2008:255). The meaning of concepts and the operational definitions of variables may differ between nations or regions (Erikson 1966:xi), so the comparative researcher must consider how best to establish measurement equivalence (Markoff 2005:402). For example, the concept of being a *good son or daughter* refers to a much broader range of behaviors in China than in most Western countries (Ho 1996). Rates of physical disability cannot be compared between nations because standard definitions are lacking (Martin and Kinsella 1995:364–365). Individuals in different cultures may respond differently to the same questions (Martin and Kinsella 1995:385). Alternatively, different measures may have been used for the same concepts in different nations, and the equivalence of these measures may be unknown (van de Vijver and Leung 1997:9). The value of statistics for particular geographic units such as counties in the United States may vary over time simply because of changes in the boundaries of these units (Walters et al. 1997). Such possibilities should be considered, and any available opportunity should be taken to test for their effects.

Exhibit 15.7 Average Percentage of Voters Who Participated in Presidential or Parliamentary Elections, 1945–1998*

Country	Vote %	Country	Vote %
Italy	92.5	St. Kitts and Nevis	58.1
Cambodia	90.5	Morocco	57.6
Seychelles	96.1	Cameroon	56.3
Iceland	89.5	Paraguay	56.0
Indonesia	88.3	Bangladesh	56.0
New Zealand	86.2	Estonia	56.0
Uzbekistan	86.2	Gambia	55.8
Albania	85.3	Honduras	55.3
Austria	85.1	Russia	55.0
Belgium	84.9	Panama	53.4
Czech	84.8	Poland	52.3
Netherlands	84.8	Uganda	50.6
Australia	84.4	Antigua and Barbuda	50.2
Denmark	83.6	Burma/Myanmar	50.0
Sweden	83.5	Switzerland	49.3
Mauritius	82.8	USA	48.3
Portugal	82.4	Mexico	48.1
Mongolia	82.3	Peru	48.0
Tuvalu	81.9	Brazil	47.9
Western Samoa	81.9	Nigeria	47.6
Andorra	81.3	Thailand	47.4
Germany	80.9	Sierra Leone	46.8
Slovenia	80.6	Botswana	46.5
Aruba	80.4	Chile	45.9
Namibia	80.4	Senegal	45.6
Greece	80.3	Ecuador	44.7
Guyana	80.3	El Salvador	44.3
Israel	80.0	Haiti	42.9
Kuwait	79.6	Ghana	42.4
Norway	79.5	Pakistan	41.8
San Marino	79.1	Zambia	40.5
Finland	79.0	Burkina Faso	38.3
Suriname	77.7	Nauru	37.3
Malta	77.6	Yemen	36.8
Bulgaria	77.5	Colombia	36.2
Romania	77.2	Niger	35.6

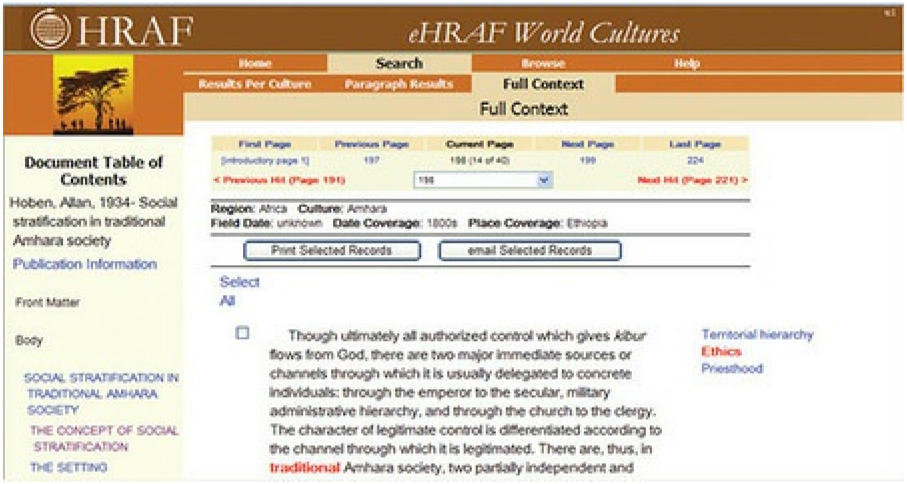
*Based on entire voting-age population in countries that held at least two elections during these years. Only countries with highest and lowest averages are shown.

Source: Reproduced by permission of International IDEA from “Turnout in the World—Country by Country Performance (1945–1998).” From Voter Turnout: A Global Survey (http://www.int/vt/survey/voter_turnout_pop2-2.cfm) © International Institute for Democracy and Electoral Assistance.

Qualitative data can also be used as a primary tool for comparative research. The Human Relations Area Files (HRAF) Collection of Ethnography provides an extraordinary resource for qualitative comparative cross-sectional research (and, to a lesser extent, for qualitative comparative historical research) (Ember and Ember 2011). The HRAF was founded in 1949 as a corporation designed to facilitate cross-cultural research. The HRAF ethnography collection now contains more than 1 million pages of material from publications and other reports from about 400 different cultural, ethnic, religious, and national groups all over the world. The information is indexed by topic, in 710 categories, and now made available electronically (if your school pays to maintain access to the HRAF). Exhibit 15.8 is an example of a page from an HRAF document that has been indexed for easy retrieval.

Most of the significant literature published on the chosen groups is included in the HRAF and used to prepare a standard summary about the group. Researchers can use these summaries and systematic searches for specific index terms to answer many questions about other social groups with the HRAF files, such as “What percentage of the world’s societies practice polygyny?” and “Does punitive child training affect the frequency of warfare?” (Ember and Ember 2011).

Exhibit 15.8 HRAF-Indexed Document


HRAF
eHRAF World Cultures
Home
Search
Browse
Help
Results Per Culture
Paragraph Results
Full Context
Full Context
First Page
Previous Page
Current Page
Next Page
Last Page
[introductory page 1]
197
198 (14 of 40)
199
224
< Previous Hit (Page 191)
198
Next Hit (Page 221) >
Print Selected Records
email Selected Records
Select All
Document Table of Contents
Publication Information
Front Matter
Body
SOCIAL STRATIFICATION IN TRADITIONAL AMHARA SOCIETY
THE CONCEPT OF SOCIAL STRATIFICATION
THE SETTING
Territorial hierarchy
Ethics
Priesthood

Hoben, Allan, 1934- Social stratification in traditional Amhara society

Region: Africa Culture: Amhara
Field Date: unknown Date Coverage: 1800s Place Coverage: Ethiopia

☐ Though ultimately all authorized control which gives *Abbur* flows from God, there are two major immediate sources or channels through which it is usually delegated to concrete individuals: through the emperor to the secular, military administrative hierarchy, and through the church to the clergy. The character of legitimate control is differentiated according to the channel through which it is legitimated. There are, thus, in traditional Amhara society, two partially independent and

Careers and Research

Ruth Westby, MA

For Ruth Westby, research—particularly public health research—means the chance to make new discoveries that affect people's lives by improving community health. She has studied how programs for disadvantaged and underserved groups are implemented and whether they have meaningful health impacts.

Westby was inspired to pursue a career in clinical research after her father died from cancer shortly after she received her BA from Emory University. After a few years of working with sick individuals on clinical trials, she decided to focus on public health so that she could look toward *preventing* disease. She sought out skill-based research courses and then internships that would help her use those skills as a graduate student. One such internship, at the Centers for Disease Control and Prevention, led to coauthored journal articles and a presentation at a large conference. In this way, Westby was exposed to opportunities that cemented her passion for public health research and provided a job in which every day at work is different and evokes a sense of pride.

Westby's research job also has kept her learning new research methods. She has already been exposed to systematic literature reviews, secondary data analyses, quantitative and qualitative data collection and analyses, and program evaluation. She finds program evaluation particularly rewarding, as she studies how programs are implemented and whether they have meaningful health impacts on disadvantaged populations.

If she could give current students advice, it would be to take advantage of mentors, faculty members, and anyone who is willing to help you learn:

> I've seen first-hand the advantages of getting to know faculty members on a personal level, networking and interning at institutions where I might want to work later, and using new research skills outside of class. Doing all of these things taught me so much more than if I had just attended lectures and read my textbooks. By the time I graduated from graduate school, I felt much more competent and set up for success than after college. In the long run, those relationships and experiences will mean just as much, if not more, than your GPA or course schedule.

Comparative Historical Research

The combination of historical analysis with comparisons between nations or other units often leads to the most interesting results. Historical social scientists may use comparisons between cases "to highlight the particular features of each case" (Skocpol 1984:370) or to identify general historical patterns across nations. A study of processes within one nation may therefore be misleading if important processes within the nation have been influenced by social processes that operate across national boundaries (Markoff 2005:403). For example, comparisons between nations may reveal that differences in political systems are much more important than voluntary decisions by individual political leaders (Rueschemeyer et al. 1992:31–36).

Comparative historical research can also help identify the causal processes at work within the nations or other entities (Lipset 1968:34; Skocpol 1984:374–386). Comparative historical research can result in historically conditional theory, in which the applicability of general theoretical propositions is linked to particular historical circumstances (Paige 1999). For example, James Mahoney (2001) explained the differential success of liberalizing reforms in Central American countries as the result of the particular ways in which liberal elites tried to implement the reforms. As summarized by Lange (2013:77–78), Guatemala and El Salvador tried to implement the reforms quickly by developing a militarized state, whereas Costa Rica used a gradual approach that could be implemented by a democratic regime. In Honduras and Nicaragua, internal pressure and external interventions stopped the liberal reforms entirely. These early events set political processes in these countries on different trajectories for many decades. In a similar way, Fallon, Swiss, and Viterna (2012) identified the influences shaping the extent of women's legislative representation in democratizing countries.

The great classical sociologists also used comparative methods, although their approach was less systematic. For example, Max Weber's (Bendix 1962:268) comparative sociology of religions contrasted Protestantism in the West, Confucianism and Taoism in China, Hinduism and Buddhism in India, and Ancient Judaism. As Bendix (1962) explained,

> His [Weber's] aim was to delineate religious orientations that contrasted sharply with those of the West, because only then could he specify the features that were peculiar to Occidental [Western] religiosity and hence called for an explanation . . . to bring out the distinctive features of each historical phenomenon. (p. 268)

So, for example, Weber concluded that the rise of Protestantism, with its individualistic approach to faith and salvation, was an important factor in the development of capitalism.

Mitchell Duneier's (2016) new book, *Ghetto: The Invention of a Place, The History of an Idea,* uses this classical approach in historical comparative research. To understand the social origins of the concept of a ghetto and its influence on social behavior, Duneier compares the development of what came to be known as ghettos in medieval Venice, Nazi-occupied Warsaw, and then Chicago and New York's Harlem in different historical periods (Chicago in the 1940s and 1980s, and Harlem in the 1960s and 2000s). Although his primary focus is on ghettos in American cities, Duneier begins his analysis by comparing the history of 16th century Venice's Jewish ghetto with the history of the Jewish ghetto in Nazi-occupied Warsaw. In both he finds "the pernicious circular logic of the ghetto": "the consequences of ghettoization provided an apparent justification for the original condition" (p. 11).

But Duneier also identifies a striking contrast: Although daytime social circulation across the ghetto boundaries allowed a wide range of social institutions to continue in spite of religious prejudices, the barbed wire barriers around Warsaw's ghetto facilitated a complete segregation that was designed to hasten its occupants' deaths. Segregation, Duneier concludes, can be compatible "with wide variations in both control and flourishing" (p. 220). When he turns to Chicago and New York, relying primarily on the work of local black scholars, Duneier finds in earlier years evidence of more functional communities that are then decimated by growing opportunities for middle-class blacks to move to the suburbs and by increasingly punitive criminal justice policies. Duneier concludes that "the ghetto is an expression of societal power" reflecting prejudice against blacks as well as "ongoing external domination and neglect" that spans generations (pp. 222–225).

Much modern comparative historical research takes a more systematic approach to comparisons, usually focusing on a detailed sequence of events that might have influenced an outcome in the present. Some studies collect quantitative longitudinal data about a number of nations and then use these data to test hypotheses about influences on national characteristics. (Theda Skocpol [1984:375] terms this *analytic historical sociology.*) Others compare the histories or particular historical experiences of nations in a narrative form, noting similarities and differences and inferring explanations for key national events (*interpretive historical sociology* in Skocpol's terminology [1984:368]). Either quantitative or qualitative data can be used.

There are several stages for a systematic, qualitative, comparative historical study (Ragin 1987:44–52; Rueschemeyer et al. 1992:36–39):

1. Specify a theoretical framework and identify key concepts or events that should be examined to explain a phenomenon.
2. Select cases (such as nations) that vary in terms of the key concepts or events.
3. Identify similarities and differences between the cases in these key concepts or events and the outcome to be explained.
4. Propose a causal explanation for the historical outcome and check it against the

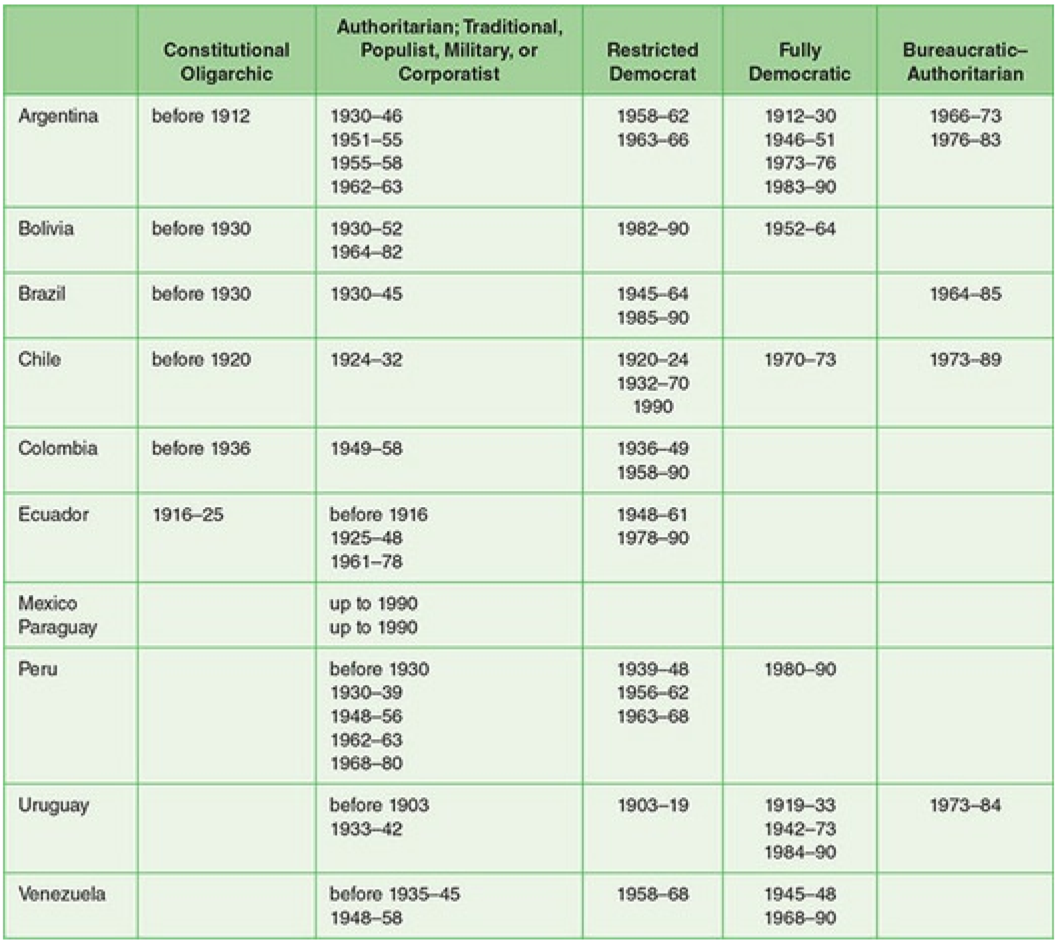
> features of each case. The criterion of success in this method is to explain the outcome for each case, without allowing deviations from the proposed causal pattern.

Dietrich Rueschemeyer et al. (1992) used a comparative historical method to explain why some nations in Latin America (excluding Central America) developed democratic politics, whereas others became authoritarian or bureaucratic–authoritarian states. First, Rueschemeyer et al. developed a theoretical framework that gave key attention to the power of social classes, state (government) power, and the interaction between social classes and the government. The researchers then classified the political regimes in each nation over time (see Exhibit 15.9). Next, they noted how each nation varied over time relative to the variables they had identified as potentially important for successful democratization.

Their analysis identified several conditions for initial democratization: consolidation of state power (ending overt challenges to state authority), expansion of the export economy (reducing conflicts over resources), industrialization (increasing the size and interaction of middle and working classes), and some agent of political articulation of the subordinate classes (which could be the state, political parties, or mass movements). Historical variation in these conditions was then examined in detail.

When geographic units such as nations are sampled for comparative purposes, it is assumed that the nations are independent of each other in the variables examined. Each nation can then be treated as a separate case for identifying possible chains of causes and effects. However, in a very interdependent world, this assumption may be misplaced—nations may develop as they do because of how other nations are developing (and the same can be said of cities and other units). As a result, comparing the particular histories of different nations may overlook the influence of global culture, international organizations, or economic dependency—just the type of influence identified in Frank et al.'s study of environmental protectionism (Skocpol 1984:384; compare Chase-Dunn and Hall 1993). These common international influences may cause the same pattern of changes to emerge in different nations; looking within the history of these nations for the explanatory influences would lead to spurious conclusions (de Vaus 2008:258). The possibility of such complex interrelations should always be considered when evaluating the plausibility of a causal argument based on a comparison between two apparently independent cases (Jervis 1996).

Exhibit 15.9 Classification of Regimes Over Time



	Constitutional Oligarchic	Authoritarian; Traditional, Populist, Military, or Corporatist	Restricted Democrat	Fully Democratic	Bureaucratic–Authoritarian
Argentina	before 1912	1930–46 1951–55 1955–58 1962–63	1958–62 1963–66	1912–30 1946–51 1973–76 1983–90	1966–73 1976–83
Bolivia	before 1930	1930–52 1964–82	1982–90	1952–64	
Brazil	before 1930	1930–45	1945–64 1985–90		1964–85
Chile	before 1920	1924–32	1920–24 1932–70 1990	1970–73	1973–89
Colombia	before 1936	1949–58	1936–49 1958–90		
Ecuador	1916–25	before 1916 1925–48 1961–78	1948–61 1978–90		
Mexico Paraguay		up to 1990 up to 1990			
Peru		before 1930 1930–39 1948–56 1962–63 1968–80	1939–48 1956–62 1963–68	1980–90	
Uruguay		before 1903 1933–42	1903–19	1919–33 1942–73 1984–90	1973–84
Venezuela		before 1935–45 1948–58	1958–68	1945–48 1968–90	

Source: Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. Capitalist Development and Democracy. Reprinted with permission from the University of Chicago Press.

Comparative Case Study Designs

Some comparative researchers use a systematic method for identifying causes that owes its origins to the English philosopher John Stuart Mill (1872). One approach that Mill developed was called the method of agreement. The core of this approach is the comparison of nations (*cases*) for similarities and differences on potential causal variables and the phenomenon to be explained. As comparative historian Skocpol (1979:36) explains, researchers who use this method should "try to establish that several cases having in common the phenomenon one is trying to explain also have in common a set of causal factors, although they vary in other ways that might have seemed causally relevant."

For example, suppose three countries that have all developed democratic political systems are compared through four socioeconomic variables hypothesized by different theories to influence democratization (see Exhibit 15.10). If the countries differ in three of the variables but are similar in the fourth, this is evidence that the fourth variable influences democratization. In Exhibit 15.10, the method of agreement would lead the analyst to conclude that an expanding middle class was a cause of the democratization experienced in all three countries. The focus of the method of agreement is actually on identifying a similarity between cases that differ in many respects, so this approach is also called the *most different case studies* method.

The second approach Mill developed was the method of difference. Again, in the words of comparative historian Skocpol (1979),

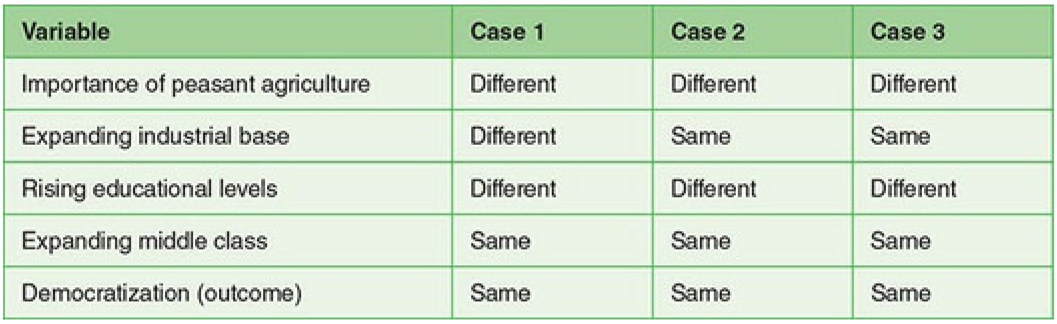
> One can contrast the cases in which the phenomenon to be explained and the hypothesized causes are present to other cases in which the phenomenon and the causes are both absent, but which are otherwise as similar as possible to the positive cases. (p. 36)

The method of difference approach is represented in Exhibit 15.11. In this example, "moderate income disparities" are taken to be the cause of democratization because the country that didn't democratize differs in this respect from the country that did democratize. These two countries are similar with respect to other potential influences on democratization. The argument could be improved by adding more positive and negative cases. The focus of the method of difference is actually on identifying a difference between cases that are similar in other respects, so this approach is also called the *most similar case studies* method.

Method of agreement: A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that agree on an outcome variable also agree on the value of the variable hypothesized to have a causal effect, although they differ on other variables.

Method of difference: A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that differ on an outcome variable also differ on the value of the variable hypothesized to have a causal effect, although they agree on other variables.

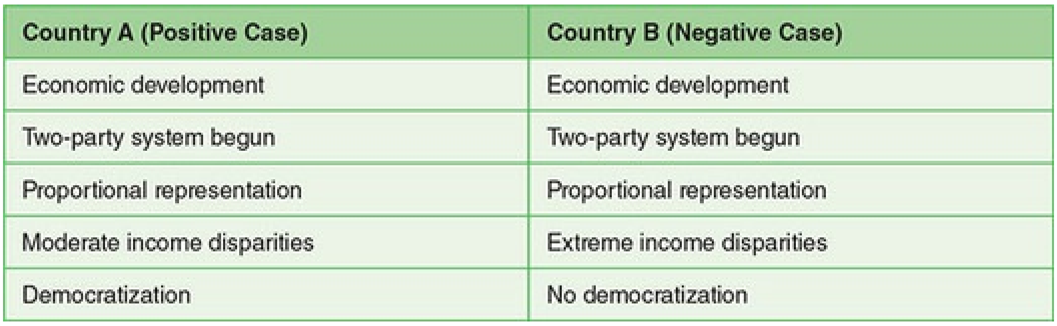
Exhibit 15.10 John Stuart Mill's Method of Agreement (Hypothetical Cases and Variables)



Variable	**Case 1**	**Case 2**	**Case 3**
Importance of peasant agriculture	Different	Different	Different
Expanding industrial base	Different	Same	Same
Rising educational levels	Different	Different	Different
Expanding middle class	Same	Same	Same
Democratization (outcome)	Same	Same	Same

Source: Adapted from Skocpol (1984:379).

Exhibit 15.11 John Stuart Mill's Method of Difference (Hypothetical Cases and Variables)



Country A (Positive Case)	**Country B (Negative Case)**
Economic development	Economic development
Two-party system begun	Two-party system begun
Proportional representation	Proportional representation
Moderate income disparities	Extreme income disparities
Democratization	No democratization

Source: Adapted from Skocpol and Somers (1979:80).

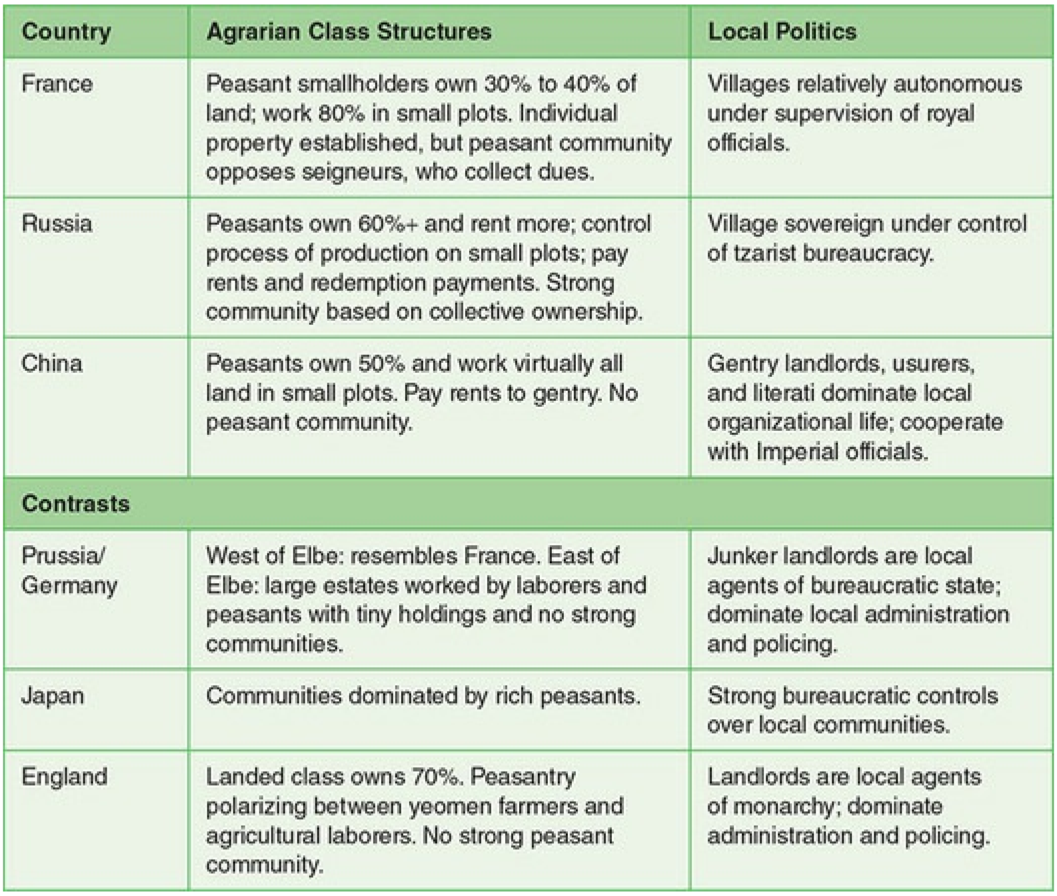
The method of agreement and method of difference approaches can also be combined, "by using at once several positive cases along with suitable negative cases as contrasts" (Skocpol 1979:37). This is the approach that Skocpol (1979) used in her classic book about the French, Russian, and Chinese revolutions, *States and Social Revolutions.* Exhibit 15.12 summarizes part of her argument about the conditions for peasant insurrections, based on a careful historical review. In this exhibit, Skocpol (1979:156) shows how the three countries that experienced revolutions (France, Russia, and China) tended to have more independent peasants and more autonomy in local politics than did three contrasting countries (Prussia/Germany, Japan, and England) that did not experience social revolutions.

Cautions for Comparative Methods

Of course, ambitious methods that compare different countries face many complications. The features of the cases selected for comparison have a large impact on the researcher's ability to identify influences. Cases should be chosen for their difference on key factors hypothesized to influence the outcome of interest and their similarity on other, possibly confounding, factors (Skocpol 1984:383). For example, to understand how industrialization influences democracy, you would need to select cases for comparison that differ in industrialization, so that you could then see if they differ in democratization (King, Keohane, and Verba 1994:148–152). Nonetheless, relying on just a small number of cases for comparisons introduces uncertainty into the conclusions (de Vaus 2008:256).

And what determines whether cases are similar and different in certain respects? In many comparative analyses, the values of continuous variables are dichotomized. For example, nations may be coded as *democratic* or *not democratic* or as having *experienced revolution* or *not experienced revolution.* The methods of agreement and difference that I have just introduced presume these types of binary (dichotomous) distinctions. However, variation in the social world often involves degrees of difference, rather than all or none distinctions (de Vaus 2008:255). Some countries may be partially democratic and some countries may have experienced a limited revolution. At the individual level, you know that distinctions such as *rich* and *poor* or *religious* and *not religious* reflect differences on underlying continua of wealth and religiosity. So the use of dichotomous distinctions in comparative analyses introduces an imprecise and somewhat arbitrary element into the analysis (Lieberson 1991). For some comparisons, however, qualitative distinctions such as *simple majority rule* or *unanimity required* may capture the important differences between cases better than quantitative distinctions. We don't want to simply ignore important categorical considerations such as this in favor of *degree of majority rule* or some other underlying variable (King et al. 1994:158–163). Careful discussion of the bases for making distinctions is an important first step in any comparative historical research (also see Ragin 2000).

Exhibit 15.12 Methods of Agreement and Difference Combined: Conditions for Peasant Insurrections



Country	Agrarian Class Structures	Local Politics
France	Peasant smallholders own 30% to 40% of land; work 80% in small plots. Individual property established, but peasant community opposes seigneurs, who collect dues.	Villages relatively autonomous under supervision of royal officials.
Russia	Peasants own 60%+ and rent more; control process of production on small plots; pay rents and redemption payments. Strong community based on collective ownership.	Village sovereign under control of tzarist bureaucracy.
China	Peasants own 50% and work virtually all land in small plots. Pay rents to gentry. No peasant community.	Gentry landlords, usurers, and literati dominate local organizational life; cooperate with Imperial officials.
Contrasts		
Prussia/ Germany	West of Elbe: resembles France. East of Elbe: large estates worked by laborers and peasants with tiny holdings and no strong communities.	Junker landlords are local agents of bureaucratic state; dominate local administration and policing.
Japan	Communities dominated by rich peasants.	Strong bureaucratic controls over local communities.
England	Landed class owns 70%. Peasantry polarizing between yeomen farmers and agricultural laborers. No strong peasant community.	Landlords are local agents of monarchy; dominate administration and policing.

Source: Skocpol (1979).

The focus on comparisons between nations may itself be a mistake for some analyses. National boundaries often do not correspond to key cultural differences, so comparing subregions within countries or larger cultural units that span multiple countries may make more sense for some analyses (de Vaus 2008:258). Comparing countries that have fractured along cultural or religious divides simply by average characteristics would obscure many important social phenomena.

With cautions such as these in mind, the combination of historical and comparative methods allows for rich descriptions of social and political processes in different nations or regions as well as for causal inferences that reflect a systematic, defensible weighing of the evidence. Data of increasingly good quality are available on a rapidly expanding number of nations, creating many opportunities for comparative research. We cannot expect one study comparing the histories of a few nations to control adequately for every plausible alternative causal influence, but repeated investigations can refine our understanding and lead to increasingly accurate causal conclusions (King et al. 1994:33).

Demographic Analysis

The social processes that are the focus of historical and comparative research are often reflected in and influenced by changes in the makeup of the population being studied. For example, the plummeting birthrates in European countries will influence the politics of immigration in those countries, their living standards, the character of neighborhoods, and national productivity (Bruni 2002). Demography is the field that studies these dynamics. Demography is the statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time. Demographers explain population change through five processes: (1) fertility, (2) mortality, (3) marriage, (4) migration, and (5) social mobility (Bogue 1969:1).

Demography: The statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time.

Demographers obtain data from a census of the population (see Chapter 5) and from registries—records of events such as births, deaths, migrations, marriages, divorces, diseases, and employment (Anderton, Barrett, and Bogue 1997:54–79; Baum 1993), then compute various statistics from these data to facilitate description and analysis (Wunsch and Termote 1978). To use these data, you need to understand how they are calculated and the questions they answer. Four concepts are key to understanding and using demographic methods: population change, standardization of population numbers, the demographic bookkeeping equation, and population composition.

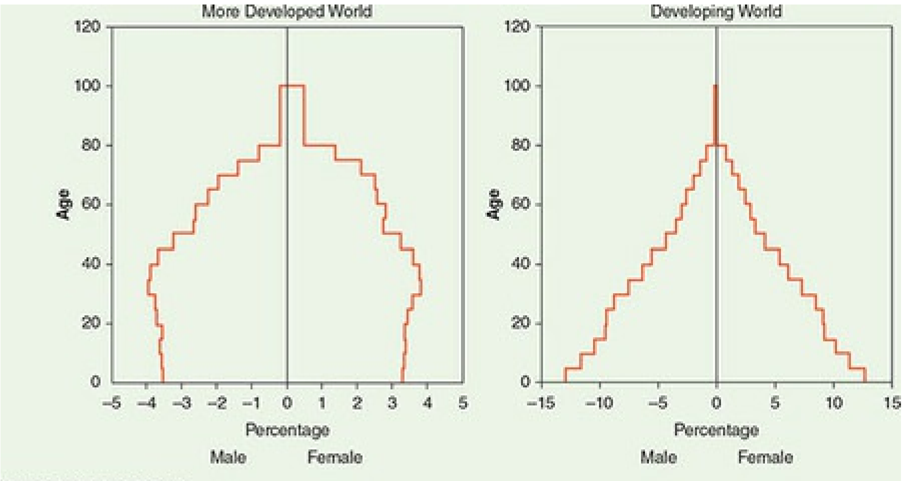
Population change is a central concept in demography. The absolute population change is calculated simply as the difference between the population size in one census minus the population size in an earlier census. This measure of absolute change is of little value, however, because it does not consider the total size of the population that was changing (Bogue 1969:32–43). A better measure is the *intercensal percent change,* which is the absolute change in population between two censuses divided by the population size in the earlier census (and multiplied by 100 to obtain a percentage). With the percent change statistic, we can meaningfully compare the growth in two or more nations that differ markedly in size (as long as the intercensal interval does not vary between the nations) (White 1993:1–2).

Standardization of population numbers, as with the calculation of intercensal percent change, is a key concern of demographic methods (Gill, Glazer, and Thernstrom 1992:478–482; Rele 1993). To make meaningful comparisons between nations and over time, numbers that describe most demographic events must be adjusted for the size of the population at risk for the event. For example, the fertility rate is calculated as the ratio of the number of births to women of childbearing age to the total number of women in this age range (multiplied by 1,000). Unless we make such adjustments, we will not know if a nation with a much higher number of births or deaths in relation to its total population size simply has more women in the appropriate age range or has more births per "eligible" woman.

The *demographic bookkeeping* (or *balancing*) *equation* is used to identify the four components of population growth during a time interval (P_2P_1): births (B), deaths (D), and in-migration (M_i) and out-migration (M_o). The equation is written as follows: $P_2 = P_1 + (B - D) + (M_i - M_o)$. That is, population at a given point in time is equal to the population at an earlier time plus the excess of births over deaths during the interval and the excess of in-migration over out-migration (White 1993:1–4). Whenever you see population size or change statistics used in a comparative analysis, you will want to ask yourself whether it is also important to know which component in the equation was responsible for the change over time or for the difference between countries (White 1993:1–4).

Population composition refers to a description of a population in terms of basic characteristics such as age, race, sex, or marital status (White 1993:1–7). Descriptions of population composition at different times or in different nations can be essential for understanding social dynamics identified in historical and comparative research. For example, Exhibit 15.13 compares the composition of the population in more developed and developing regions of the world by age and sex in 1995, using United Nations data. By comparing these *population pyramids,* we see that children constitute a much greater proportion of the population in less developed regions. The more developed regions' population pyramid also shows the greater proportion of women at older ages and the post–World War II baby boom bulge in the population.

Exhibit 15.13 Population Pyramids for More Developed and Developing Regions of the World: 1995*



*Inconsistent date in original.

Source: Bogue, Donald J., Eduardo E. Arriaga, and Douglas L. Anderton. 1993. Readings in Population Research Methodology, vol. 1, Basic Tools. Chicago: Social Development Center, for the United Nations Populations Fund.

Demographic analysis can be an important component of historical research (Bean, Mineau, and Anderton 1990), but problems of data quality must be evaluated carefully (Vaessen 1993). The hard work that can be required to develop demographic data from evidence that is hundreds of years old does not always result in worthwhile information. The numbers of people for which data are available in particular areas may be too small for statistical analysis, data that are easily available (e.g., a list of villages in an area) may not provide the information that is important (e.g., population size), and lack of information on the original data collection procedures may prevent assessment of data quality (Hollingsworth 1972:77).

Content Analysis

How are medical doctors regarded in U.S. culture? Do newspapers use the term *schizophrenia* in a way that reflects what this serious mental illness actually involves? Does the portrayal of men and women in video games reinforce gender stereotypes? Are the body images of male and female college students related to their experiences with romantic love? If you are concerned with understanding culture, attitudes toward mental illness, or gender roles, you'll probably find these to be important research questions. You now know that you could probably find data about each of these issues for a secondary data analysis, but in this section, I would like to introduce procedures for analyzing a different type of data that awaits the enterprising social researcher. Content analysis is "the systematic, objective, quantitative analysis of message characteristics" and is a method particularly well suited to the study of popular culture and many other issues concerning human communication (Neuendorf 2002:1).

Content analysis can be used in historical and comparative research but is also useful in studies of communication with any type of media. Therefore, like most historical and comparative methods, content analysis can be called an *unobtrusive method* that does not need to involve interacting with live people. Content analysis methods usually begin with text, speech broadcasts, or visual images. The content analyst develops procedures for coding various aspects of the textual, aural (spoken), or visual material and then analyzes this coded content.

The goal of content analysis is to develop inferences from human communication in any of its forms, including books, articles, magazines, songs, films, and speeches (Weber 1990:9). You can think of content analysis as a "survey" of some documents or other records of communication—a survey with fixed-choice responses that produce quantitative data. This method was first applied to the study of newspaper and film content and then developed systematically for the analysis of Nazi propaganda broadcasts in World War II. Since then, content analysis has been used to study historical documents, records of speeches, and other "voices from the past" as well as media of all sorts (Neuendorf 2002:31–37). The same techniques can now be used to analyze blog sites, wikis, and other text posted on the Internet (Gaiser and Schreiner 2009:81–90). Content analysis techniques are also used to analyze responses to open-ended survey questions.

Content analysis bears some similarities to qualitative data analysis because it involves coding and categorizing text and identifying relationships between constructs identified in the text. However, because it usually is conceived as a quantitative procedure, content analysis overlaps with qualitative data analysis only at the margins—the points where qualitative analysis takes on quantitative features or where content analysis focuses on qualitative features of the text. This distinction becomes fuzzy, however, because content analysis techniques can be used with all forms of messages, including visual images, sounds, and interaction patterns, as well as written text (Neuendorf 2002:24–25).

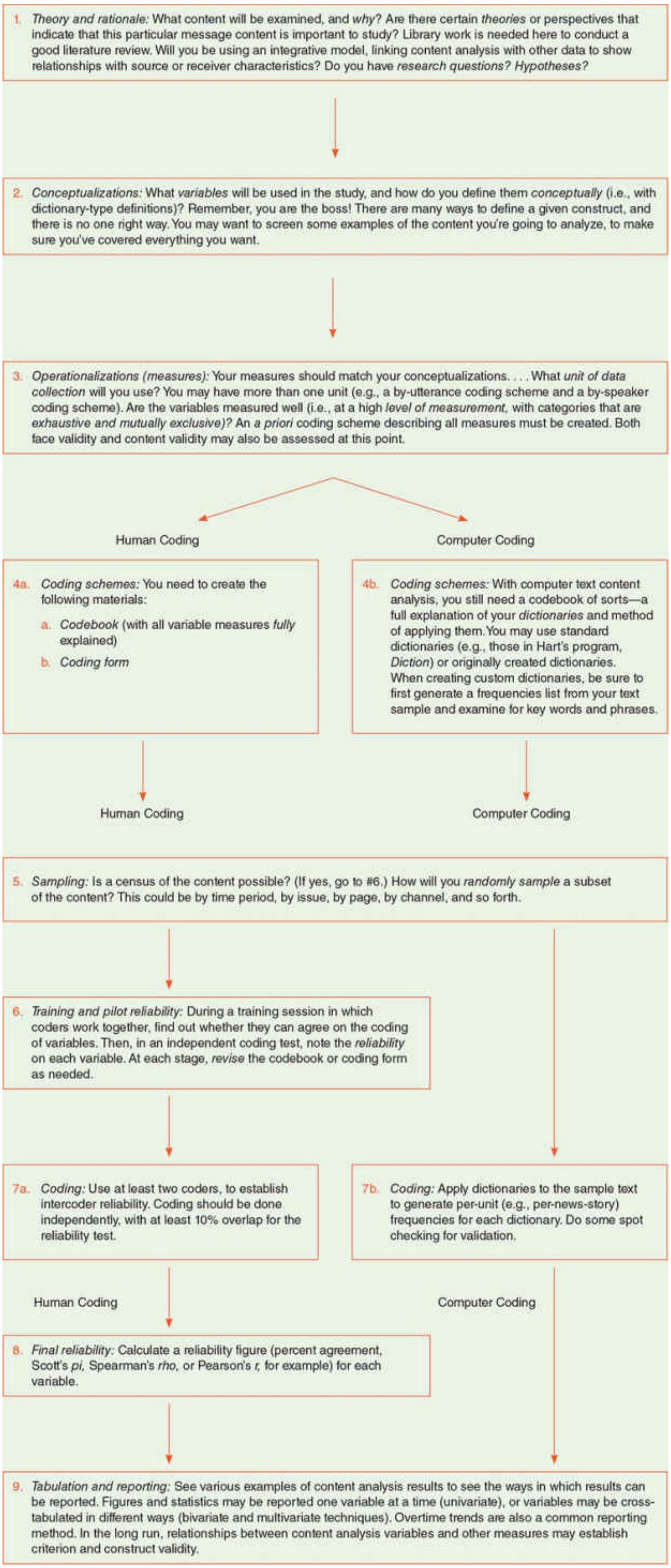
The various steps in a content analysis are represented in the flowchart in Exhibit 15.14. Note that the steps are comparable to the procedures in quantitative survey research. Use this flowchart as a checklist when you design or critique a content analysis project.

Kimberly Neuendorf's (2002:3) analysis of medical prime-time network television programming introduces the potential of content analysis. As Exhibit 15.15 shows, medical programming has been dominated by noncomedy shows, but there have been two significant periods of comedy medical shows—during the 1970s and early 1980s and then again in the early 1990s. It took a qualitative analysis of medical show content to reveal that the 1960s shows represented a very distinct "physician-as-God" era, which shifted to a more human view of the medical profession in the 1970s and 1980s. This era has been followed, in turn, by a mixed period that has had no dominant theme.

Content analysis is useful for investigating many questions about the social world. To illustrate its diverse range of applications, I will use in the next sections Neuendorf's (2002) analysis of TV programming; Matthias A. Gerth and Gabriele Siegert's (2012) analysis of news coverage of an immigrant naturalization campaign; Kenneth Duckworth's, John Halpern's, Chris Gillespie's, and my (2003) analysis of newspaper articles; Karen Dill and Kathryn Thill's (2007) analysis of video game characters; and Suman Ambwani and Jaine Strauss's (2007) analysis of student responses to open-ended survey questions. These examples will demonstrate that the units that are "surveyed" in a content analysis can range from newspapers, books, or TV shows to persons referred to in other communications, themes expressed in documents, or propositions made in different statements.

Content analysis: A research method for systematically analyzing and making inferences from recorded human communication, including books, articles, poems, constitutions, speeches, and songs.

Exhibit 15.14 Flowchart for the Typical Process of Content Analysis Research



1. *Theory and rationale:* What content will be examined, and *why*? Are there certain *theories* or perspectives that indicate that this particular message content is important to study? Library work is needed here to conduct a good literature review. Will you be using an integrative model, linking content analysis with other data to show relationships with source or receiver characteristics? Do you have *research questions? Hypotheses?*

2. *Conceptualizations:* What *variables* will be used in the study, and how do you define them *conceptually* (i.e., with dictionary-type definitions)? Remember, you are the boss! There are many ways to define a given construct, and there is no one right way. You may want to screen some examples of the content you're going to analyze, to make sure you've covered everything you want.

3. *Operationalizations (measures):* Your measures should match your conceptualizations. . . . What *unit of data collection* will you use? You may have more than one unit (e.g., a by-utterance coding scheme and a by-speaker coding scheme). Are the variables measured well (i.e., at a high *level of measurement,* with categories that are *exhaustive and mutually exclusive*)? An *a priori* coding scheme describing all measures must be created. Both face validity and content validity may also be assessed at this point.

4a. *Coding schemes:* You need to create the following materials:

a. *Codebook* (with all variable measures *fully* explained)

b. *Coding form*

4b. *Coding schemes:* With computer text content analysis, you still need a codebook of sorts—a full explanation of your *dictionaries* and method of applying them. You may use standard dictionaries (e.g., those in Hart's program, *Diction*) or originally created dictionaries. When creating custom dictionaries, be sure to first generate a frequencies list from your text sample and examine for key words and phrases.

5. *Sampling:* Is a census of the content possible? (If yes, go to #6.) How will you *randomly sample* a subset of the content? This could be by time period, by issue, by page, by channel, and so forth.

6. *Training and pilot reliability:* During a training session in which coders work together, find out whether they can agree on the coding of variables. Then, in an independent coding test, note the *reliability* on each variable. At each stage, *revise* the codebook or coding form as needed.

7a. *Coding:* Use at least two coders, to establish intercoder reliability. Coding should be done independently, with at least 10% overlap for the reliability test.

7b. *Coding:* Apply dictionaries to the sample text to generate per-unit (e.g., per-news-story) frequencies for each dictionary. Do some spot checking for validation.

8. *Final reliability:* Calculate a reliability figure (percent agreement, Scott's *pi,* Spearman's *rho,* or Pearson's *r,* for example) for each variable.

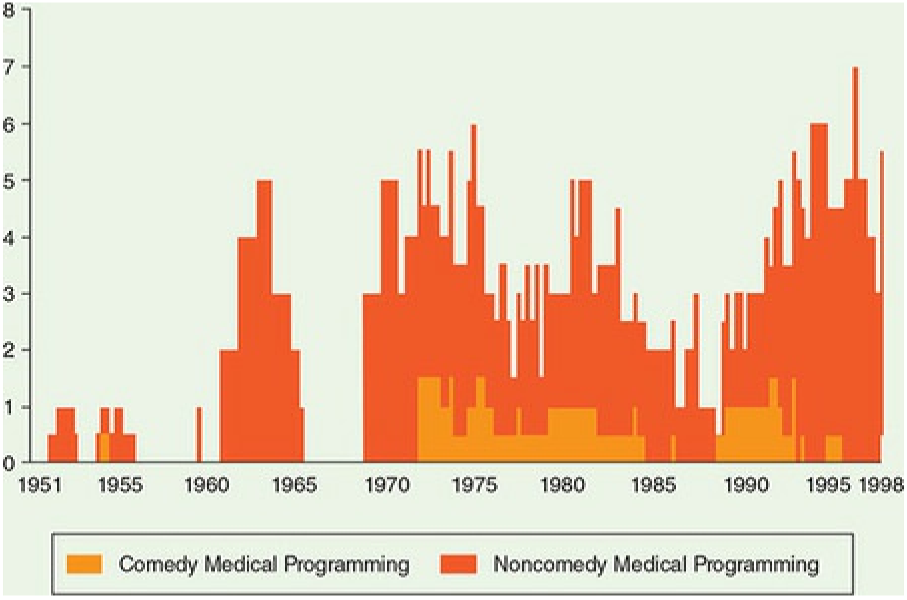
9. *Tabulation and reporting:* See various examples of content analysis results to see the ways in which results can be reported. Figures and statistics may be reported one variable at a time (univariate), or variables may be cross-tabulated in different ways (bivariate and multivariate techniques). Overtime trends are also a common reporting method. In the long run, relationships between content analysis variables and other measures may establish criterion and construct validity.

Source: Neuendorf (2002).

Identify a Population of Documents or Other Textual Sources

The population of documents or other textual sources should be selected so that it is appropriate to the research question of interest. Perhaps the population will be all newspapers published in the United States, college student newspapers, nomination speeches at political party conventions, or "state of the nation" speeches by national leaders. Books or films are also common sources for content analysis projects. Often, a comprehensive archive can provide the primary data for the analysis (Neuendorf 2002:76–77). For a fee, the LexisNexis service makes a large archive of newspapers available for analysis. For her analysis of prime-time programming since 1951, Neuendorf (2002:3–4) used a published catalog of all TV shows. For my analysis with Duckworth and others (2003:1402) of newspapers' use of the terms *schizophrenia* and *cancer,* I requested a sample of articles from the LexisNexis national newspaper archive. Gerth and Siegert (2012) focused on TV and newspaper stories during a 14-week Swiss political campaign, and Dill and Thill (2007:855–856) turned to video game magazines for their analysis of the depiction of gender roles in video games. For their analysis of gender differences in body image and romantic love, Ambwani and Strauss (2007:15) surveyed students at a small midwestern liberal arts college.

Exhibit 15.15 Medical Prime-Time Network Television Programming, 1951–1998



Source: Neuendorf (2002).

Determine the Units of Analysis

The units of analysis could be items such as newspaper articles, whole newspapers, speeches, or political conventions, or they could be more microscopic units such as words, interactions, time periods, or other bits of a communication (Neuendorf 2002:71). The content analyst has to decide what units are most appropriate to the research question and how the communication content can be broken into those units. If the units are individual issues of a newspaper, in a study of changes in news emphases, this step may be relatively easy. However, if the units are most appropriately the instances of interaction between characters in a novel or a movie, in a study of conflict patterns between different types of characters, it will require a careful process of testing to determine how to define operationally the specific units of interaction (Weber 1990:39–40).

Units of analysis varied across the five content analysis projects I have introduced. The units of analysis for Neuendorf (2002:2) were "the individual medically oriented TV program"; for Duckworth et al. (2003:1403), they were newspaper articles; for Gerth and Siegert (2012:288) they were arguments made in media stories; and for Dill and Thill (2007:856) they were images appearing in magazine articles. The units of analysis for Ambwani and Strauss (2007:15) were individual students.

Select a Sample of Units From the Population

The simplest strategy might be to select a simple random sample of documents. However, a stratified sample might be needed to ensure adequate representation of community newspapers in large and in small cities, or of weekday and Sunday papers, or of political speeches during election years and in off years (see Chapter 4) (Weber 1990:40–43). Nonrandom sampling methods have also been used in content analyses when the entire population of interest could not be determined (Neuendorf 2002:87–88).

The selected samples in our five content analysis projects were diverse. In fact, Neuendorf (2002:2) included the entire population of medically oriented TV programs between 1951 and 1998. For my content analysis with Duckworth (Duckworth et al. 2003), I had my student, Chris Gillespie, draw a stratified random sample of 1,802 articles published in the five U.S. newspapers with the highest daily circulation in 1996 to 1997 in each of the four regions identified in the LexisNexis database, as well as the two high-circulation national papers in the database, *The New York Times* and *USA Today* (pp. 1402–1403).

Because individual articles cannot be sampled directly in the LexisNexis database, a random sample of days was drawn first. All articles using the terms *schizophrenia* or *cancer* (or several variants of these terms) were then selected from the chosen newspapers on these days. Gerth and Siegert (2012:285) selected 24 different newspapers and 5 TV news programs that targeted the population for the campaign, and they then coded 3,570 arguments made in them about the campaign during its 14 weeks. Dill and Thill (2007:855–856) used all images in the current issues (as of January 2006) of the six most popular video game magazines sold on Amazon.com. Ambwani and Strauss (2007:15) used an availability sampling strategy, with 220 students from Introductory Psychology and a variety of other sources.

Design Coding Procedures for the Variables to Be Measured

Designing coding procedures requires deciding what variables to measure, using the unit of text to be coded, such as words, sentences, themes, or paragraphs. Then, the categories into which the text units are to be coded must be defined. These categories may be broad, such as *supports democracy,* or narrow, such as *supports universal suffrage.* Reading or otherwise reviewing some of the documents or other units to be coded is an essential step in thinking about variables that should be coded and in developing coding procedures. Development of clear instructions and careful training of coders are essential.

As an example, Exhibit 15.16 is a segment of the coding form that I developed for a content analysis of union literature that I collected during a mixed-methods study of union political processes (Schutt 1986). My sample was of 362 documents: all union newspapers and a stratified sample of union leaflets given to members during the years of my investigation. My coding scheme included measures of the source and target for the communication, as well as measures of concepts that my theoretical framework indicated were important in organizational development: types of goals, tactics for achieving goals, organizational structure, and forms of participation. The analysis documented a decline in concern with client issues and an increase in focus on organizational structure, which were trends that also emerged in interviews with union members.

Exhibit 15.16 Union Literature Coding Form

I. Preliminary Codes

1. Document # ______________
2. Date ______________________

    mo yr

3. Length of text __________ pp. (round up to next 1/4 page; count legal size as 1.25)
4. Literature Type
    1. General leaflet for members/employees
    2. Newspaper/Newsletter article
    3. Rep Council motions
    4. Other material for Reps, Stewards, Delegates (e.g., budget, agenda)
    5. Activity reports of officers, President's Report
    6. Technical information-filing grievances, processing forms
    7. Buying plans/Travel packages
    8. Survey Forms, Limited Circulation material (correspondence)
    9. Non-Union
    10. Other ________________________________ (specify)

4A. If newspaper article

Position

1. Headline story
2. Other front page
3. Editorial
4. Other

4B. If Rep Council motion

Sponsor

1. Union leadership
2. Office
3. Leadership faction
4. Opposition faction
5. Other

5. Literature content—Special issues
    1. First strike (1966)
    2. Second strike (1967)
    3. Collective bargaining (1977)
    4. Collective bargaining (1979)
    5. Election/campaign literature
    6. Affiliation with AFSCME/SEIU/other national union
    7. Other

II. Source and Target

6. Primary source (code in terms of those who prepared this literature for distribution).
    1. Union-newspaper (Common Sense; IUPAE News)
    2. Union-newsletter (Info and IUPAE Bulletin)
    3. Union-unsigned
    4. Union officers
    5. Union committee
    6. Union faction (the Caucus; Rank-and-Filers; Contract Action, other election slate; PLP News; Black Facts)
    7. Union members in a specific work location/office
    8. Union members—other
    9. Dept. of Public Aid/Personnel
    10. DVR/DORS
    11. Credit Union
    12. Am. Buyers' Assoc.
    13. Other nonunion
7. Secondary source (use for lit. at least in part reprinted from another source, for distribution to members)
    1. Newspaper—general circulation
    2. Literature of other unions, organizations
    3. Correspondence of union leaders
    4. Correspondence from DPA/DVR-DORS/Personnel
    5. Correspondence from national union
    6. Press release
    7. Credit Union, Am. Buyers'
    8. Other _____________________________ (specify)
    9. None
8. Primary target (the audience for which the literature is distributed)
    1. Employees—general (if mass-produced and unless otherwise stated)
    2. Employees-DVR/DORS
    3. Union members (if refers only to members or if about union elections)
    4. Union stewards, reps, delegates committee
    5. Non-unionized employees (recruitment lit, etc.)
    6. Other _____________________________ (specify)
    7. Unclear

I. Issues

A. Goal

B. Employee conditions/benefits (Circle up to 5)

1. Criteria for hiring
2. Promotion
3. Work out of Classification, Upgrading
4. Step increases
5. Cost-of-living, pay raise, overtime pay, "money"
6. Layoffs (nondisciplinary); position cuts
7. Workloads, Redeterminations, "30 for 40," GA Review
8. Office physical conditions, safety
9. Performance evaluations
10. Length of workday
11. Sick benefits/leave—holidays, insurance, illness, vacation, voting time
12. Educational leave
13. Grievances—change in procedures
14. Discrimination (race, sex, age, religion, national origin)
15. Discipline—political (union-related)
16. Discipline—performance, other
17. Procedures with clients, at work
18. Quality of work, "worthwhile jobs"—other than relations with clients

> Source: Reprinted by permission from Organization in a Changing Environment: Unionization of Welfare Employees, edited by Russell K. Schutt, the State University of New York Press © 1986, State University of New York. All rights reserved.

Developing reliable and valid coding procedures deserves special attention in a content analysis, for it is not an easy task. The meaning of words and phrases is often ambiguous. Homographs create special problems (words such as *mine* that have different meanings in different contexts), as do many phrases that have special meanings (such as *point of no return*) (Weber 1990:29–30). As a result, coding procedures cannot simply categorize and count words; text segments in which the words are embedded must also be inspected before codes are finalized. Because different coders may perceive different meanings in the same text segments, explicit coding rules are required to ensure coding consistency. Special dictionaries can be developed to keep track of how the categories of interest are defined in the study (Weber 1990:23–29).

After coding procedures are developed, their reliability should be assessed by comparing different coders' codes for the same variables. Computer programs for content analysis can enhance reliability by facilitating the consistent application of text-coding rules (Weber 1990:24–28). Validity can be assessed with a construct validation approach by determining the extent to which theoretically predicted relationships occur (see Chapter 4).

Neuendorf's (2002:2) analysis of medical programming measured two variables that did not need explicit coding rules: length of show in minutes and the year(s) the program was aired. She also coded shows as comedies or noncomedies, as well as medical or not, but she does not report the coding rules for these distinctions. We provided a detailed description of coding procedures in our analysis of newspaper articles that used the terms *schizophrenia* or *cancer* (Duckworth et al. 2003). This description also mentions our use of a computerized text-analysis program and procedures for establishing measurement reliability:

> Content coding was based on each sentence in which the key term was used. Review of the full text of an article resolved any potential ambiguities in proper assignment of codes. Key terms were coded into one of eight categories: metaphor, obituary, first person or human interest, medical news, prevention or education, incidental, medically inappropriate, and charitable foundation. Fifty-seven of the 913 articles that mentioned schizophrenia, but none of those that mentioned cancer, were too ambiguous to be coded into one of these defined categories and so were excluded from final comparisons. Coding was performed by a trained graduate assistant using QSR's NUD*IST program. In questionable cases, final decisions were made by consensus of the two psychiatrist coauthors. A

> random subsample of 100 articles was also coded by two psychiatry residents, who, although blinded to our findings, assigned the same primary codes to 95 percent of the articles. (p. 1403)

Dill and Thill (2007) used two coders and a careful training procedure for their analysis of the magazine images about video games:

> One male and one female rater, both undergraduate psychology majors, practiced on images from magazines similar to those used in the current investigation. Raters discussed these practice ratings with each other and with the first author until they showed evidence of properly applying the coding scheme for all variables. Progress was also checked part way through the coding process, as suggested by Cowan (2002). Specifically, the coding scheme was re-taught by the first author, and the two raters privately discussed discrepancies and then independently assessed their judgments about their ratings of the discrepant items. They did not resolve discrepancies, but simply reconsidered their own ratings in light of the coding scheme refresher session. Cowan (2002) reports that this practice of reevaluating ratings criteria is of particular value when coding large amounts of violent and sexual material because, as with viewers, coders suffer from desensitization effects. (p. 856)

Ambwani and Strauss (2007) also designed a careful training process to achieve acceptable levels of reliability before their raters coded the written answers to their open-ended survey questions:

> We developed a coding scheme derived from common themes in the qualitative responses; descriptions of these themes appear in the Appendix. We independently developed lists of possible coding themes by reviewing the responses and then came to a consensus regarding the most frequently emerging themes. Next, a team of four independent raters was trained in the coding scheme through oral, written, and group instruction by the first author. The raters first coded a set of hypothetical responses; once the raters reached an acceptable level of agreement on the coding for the hypothetical responses, they began to code the actual data. These strategies, recommended by Orwin (1994), were employed to reduce error in the data coding. All four independent coders were blind to the hypotheses. Each qualitative response could be coded for multiple themes. Thus, for example, a participant who described the influence of body image in mate selection and quality of sexual relations received positive codes for both *choice* and *quality of sex.* (p. 16)

Gerth and Siegert (2012) do not provide details about their coding procedures.

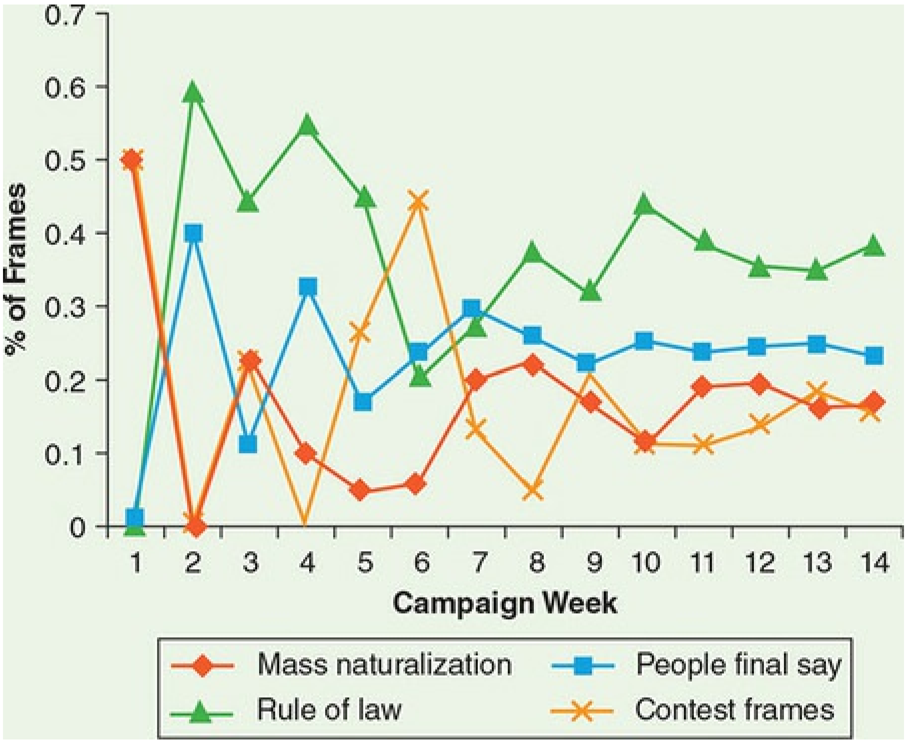
Develop Appropriate Statistical Analyses

The content analyst creates variables for analysis by counting occurrences of particular words, themes, or phrases and then tests relations between the resulting variables. These analyses could use some of the statistics introduced in Chapter 9, including frequency distributions, measures of central tendency and variation, cross-tabulations, and correlation analysis (Weber 1990:58–63). Computer-aided qualitative analysis programs, like those you learned about in Chapter 11 and like the one I selected for the preceding newspaper article analysis, can help, in many cases, to develop coding procedures and then carry out the content coding.

The simple chart that Neuendorf (2002:3) used to analyze the frequency of medical programming appears in Exhibit 15.15. My content analysis with Duckworth and others (2003) was simply a comparison of percentages showing that 28% of the articles mentioning schizophrenia used it as a metaphor, compared with only 1% of the articles mentioning cancer. We also presented examples of the text that had been coded into different categories. For example, *the nation's schizophrenic perspective on drugs* was the type of phrase coded as a metaphorical use of the term *schizophrenia* (p. 1403). Dill and Thill (2007:858) presented percentages and other statistics that showed that, among other differences, female characters were much more likely to be portrayed in sexualized ways in video game images than were male characters. Ambwani and Strauss (2007:16) used other statistics that showed that body esteem and romantic love experiences are related, particularly for women. They also examined the original written comments and found further evidence for this relationship. For example, one woman wrote, "[My current boyfriend] taught me to love my body. Now I see myself through his eyes, and I feel beautiful" (p. 17).

Gerth and Siegert (2012:288–295) use both charts and percentage distributions to test the hypotheses they posed about media attention to different political perspectives. Exhibit 15.17 shows how the use of different perspectives (or "frames") varied over the weeks of the campaign, with an emphasis on the "rule of law" being the most common way of framing the issue of naturalization.

Exhibit 15.17 Frames per Campaign Week



Source: Gerth, Matthias A. and Gabriele Siegert. 2012. "Patterns of Consistence and Constriction: How News Media Frame the Coverage of Direct Democratic Campaigns." American Behavioral Scientist 56:279–299.

The criteria for judging quantitative content analyses of text are the same standards of validity applied to data collected with other quantitative methods. We must review the sampling approach, the reliability and validity of the measures, and the controls used to strengthen any causal conclusions.

Ethical Issues in Historical and Comparative Research and Content Analysis

Analysis of historical documents, documents from other countries, or content in media does not create the potential for harm to human subjects that can be a concern when collecting primary data. It is still important to be honest and responsible in working out arrangements for data access when data must be obtained from designated officials or data archivists, but many data are available easily in libraries or on the web. Researchers in the United States who conclude that they are being denied access to public records of the federal government may be able to obtain the data by filing a Freedom of Information Act (FOIA) request. The FOIA stipulates that all persons have a right to access all federal agency records unless the records are specifically exempted (Riedel 2000:130–131). Researchers who review historical or government documents must also try to avoid embarrassing or otherwise harming named individuals or their descendants by disclosing sensitive information.

Ethical concerns are multiplied when surveys are conducted or data are collected in other countries. If the outside researcher lacks much knowledge of local norms, values, and routine activities, the potential for inadvertently harming subjects is substantial. For this reason, cross-cultural researchers should spend time learning about each of the countries in which they plan to collect primary data and strike up collaborations with researchers in those countries (Hantrais and Mangen 1996). Local advisory groups may also be formed in each country so that a broader range of opinion is solicited when key decisions must be made. Such collaboration can also be invaluable when designing instruments, collecting data, and interpreting results.

Cross-cultural researchers who use data from other societies have a particular obligation to try to understand the culture and norms of those societies before they begin secondary data analyses. It is a mistake to assume that questions asked in other languages or cultural contexts will have the same meaning as when asked in the researcher’s own language and culture, so a careful, culturally sensitive process of review by knowledgeable experts must precede measurement decisions in these projects. Ethical standards themselves may vary between nations and cultures, so cross-cultural researchers should consider collaborating with others in the places to be compared and take the time to learn about cultural practices, gender norms, and ethical standards (Ayhan 2001; Stake and Rizvi 2009:527).

Conclusions

Historical and comparative social science investigations use a variety of techniques that range from narrative histories having much in common with qualitative methods to analyses of secondary data that are in many respects like traditional survey research. Content analysis may also be used in historical and comparative research. Demographic analysis can enrich other forms of historical and comparative research by describing differences in population structure and dynamics between countries and over time. Each of these techniques can help the researchers gain new insights into processes such as democratization. They encourage intimate familiarity with the cause of development of the nations studied and thereby stimulate inductive reasoning about the interrelations between different historical events. Systematic historical and comparative techniques can be used to test deductive hypotheses concerning international differences as well as historical events.

Most historical and comparative methods encourage causal reasoning. They require the researcher to consider systematically the causal mechanism, or historical sequences of events, by which earlier events influence later outcomes. They also encourage attention to causal context, with a particular focus on the ways in which different cultures and social structures may result in different effects of other variables. Content analysis methods focus attention on key dimensions of variation in news media and other communicative texts, while demographic analysis adds another important dimension. There is much to be gained by learning and continuing to use and develop these methods.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Case-oriented research	554
Comparative historical research	553
Conjunctural research	554
Content analysis	575
Cross-sectional comparative research	553
Demography	573
Event-structure analysis	555
Historical events research	553
Historical process research	553
Holistic research	554
Method of agreement	570
Method of difference	570
Narrative explanation	554
Oral history	557
Temporal research	554
Variable-oriented research	561

Highlights

- The central insight behind historical and comparative methods is that we can improve our understanding of social processes when we make comparisons with other times and places.
- There are four basic types of historical and comparative research methods: (1) historical events research, (2) historical process research, (3) cross-sectional comparative research, and (4) comparative historical research. Historical events research and historical process research are likely to be qualitative, whereas comparative studies are often quantitative; however, research of each type may use elements of both.
- Event-structure analysis is a systematic qualitative approach to developing an idiographic causal explanation for a key event.
- Oral history provides a means of reconstructing past events. Data from other sources should be used whenever possible to evaluate the accuracy of memories.
- Qualitative historical process research uses a narrative approach to causal explanation, in which historical events are treated as part of a developing story. Narrative explanations are temporal, holistic, and conjunctural.
- Comparative methods may be cross-sectional, such as when variation between country characteristics is compared, or historical, in which developmental patterns are compared between countries.
- Methodological challenges for comparative and historical research include missing data, variation in the meaning of words and phrases and in the boundaries of geographic units across historical periods and between cultures, bias or inaccuracy of historical documents, lack of measurement equivalence, the need for detailed knowledge of the cases chosen, a limited number of cases, case selection on an availability basis, reliance on dichotomous categorization of cases, and interdependence of cases selected.
- Central concepts for demographic research are population change, standardization of population numbers, the demographic bookkeeping equation, and population composition.
- Content analysis is a tool for systematic quantitative analysis of documents and other textual data. It

requires careful testing and control of coding procedures to achieve reliable measures.

Discussion Questions

1. Review the differences between case-oriented, historically specific, inductive explanations and those that are more variable oriented, theoretically general, and deductive. List several arguments for and against each approach. Which is more appealing to you and why?
2. What historical events have had a major influence on social patterns in the nation? The possible answers are too numerous to list, ranging from any of the wars to major internal political conflicts, economic booms and busts, scientific discoveries, and legal changes. Pick one such event in your own nation for this exercise. Find one historical book on this event and list the sources of evidence used. What additional evidence would you suggest for a social science investigation of the event?
3. Consider the comparative historical research by Rueschemeyer et al. (1992) on democratic politics in Latin America. What does comparison between nations add to the researcher's ability to develop causal explanations?
4. Kathleen Fallon, Liam Swiss, and Jocelyn Vitema (2012) developed a *nomothetic* causal explanation of variation in representation of women in national legislatures over time, whereas Griffin's (1993) explanation of a lynching can be termed *idiographic.* Discuss the similarities and differences between these types of causal explanation. Use these two studies to illustrate the strengths and weaknesses of each.
5. Select a major historical event or process, such as the Great Depression, the civil rights movement, the Vietnam War, or Brexit. Why do you think this event happened? Now, select one of the four major types of historical and comparative methods that you think could be used to test your explanation. Why did you choose this method? What type of evidence would support your proposed explanation? What problems might you face in using this method to test your explanation?
6. Consider the media that you pay attention to in your social world. How could you design a content analysis of the messages conveyed by these media? What research questions could you help to answer by adding a comparison to another region or country to this content analysis?

Practice Exercises

1. The journals *Social Science History* and *Journal of Social History* report many studies of historical processes. Select one article from a recent journal issue about a historical process used to explain some event or other outcome. Summarize the author's explanation. Identify any features of the explanation that are temporal, holistic, and conjunctural. Prepare a chronology of the important historical events in that process. Do you agree with the author's causal conclusions? What additional evidence would strengthen the author's argument?
2. Exhibit 15.18 identifies voting procedures and the level of turnout in one election for 10 countries. Do voting procedures appear to influence turnout in these countries? To answer this question using Mill's methods, you will first have to decide how to dichotomize the values of variables that have more than two values (postal voting, proxy voting, and turnout). You must also decide what to do about missing values. Apply Mill's method of agreement to the pattern in the table. Do any variables emerge as likely causes? What additional information would you like to have for your causal analysis?
3. Using your library's government documents collection or the U.S. Census site on the web, select one report by the U.S. Census Bureau about the population of the United States or some segment of it. Outline the report and list all the tables included in it. Summarize the report in two paragraphs. Suggest a historical or comparative study for which this report would be useful.
4. Find a magazine or newspaper report on a demographic issue, such as population change or migration. Explain how one of the key demographic concepts could be used or was used to improve understanding of this issue.
5. Review the interactive exercises on the book's study site, at edge.sagepub.com/schutt9e, for a lesson that will help you master the terms used in historical and comparative research.
6. Select an article from the study site that used a historical or comparative design. Which specific type of design was used? What were the advantages of this design for answering the research question posed?

Exhibit 15.18 Voting Procedures in 10 Countries

	Voting Age	**Number of Days Polling Booth Open**	**Voting Day on Work Day or Rest Day**	**Postal Voting**	**Proxy Voting**	**Constituency Transfer**	**Advance Voting**	**Voter Turnout (in %)**	**Year (P = presidential, L = legislative election)**
Switzerland	20	2	Rest day	Automatic for armed forces, otherwise by application 4 days before voting	Varies by canton	No	No	46	1991L
Taiwan	20	1	Rest day					72	1992L
Thailand	20	1	Rest day	No				62	1995L
Turkey	20	1	Rest day	No	No	Special polling stations at border posts for citizens residing abroad	No	80	1991L
Ukraine	18	1	Rest day					71.6	1994P
United Kingdom	18	1	Work day	On application	On application	No	No	77.8	1992L
United States	18	1	Work day	By application; rules vary across states	In some states for blind and disabled	No		51.5	1992P
Uruguay	18	1	Rest day	No	No	No		89.4	1994P
Venezuela	18	1	Rest day	No	Assisted voting for blind and disabled	No	No	60	1993P
Zambia		1	Work day		No			50	1991P

Source: LeDuc et al. (1996:19, Figure 1.3).

Ethics Questions

1. Oral historians can uncover disturbing facts about the past. What if a researcher were conducting an oral history project such as the Depression-era Federal Writers' Project and learned from an interviewee about his previously undisclosed involvement in a predatory sex crime many years ago? Should the researcher report what he learned to a government attorney who might decide to bring criminal charges? What about informing the victim and/or her surviving relatives? Would it matter if the statute of limitations had expired, so that the offender could not be prosecuted any longer? After Boston College researchers interviewed former participants in "The Troubles," the police in Northern Ireland subsequently demanded and received some of the transcripts and used them in their investigations (Cullen 2014). Should the oral history project not have been conducted?
2. In this chapter's ethics section, I recommended that researchers who conduct research in other cultures form an advisory group of local residents to provide insight into local customs and beliefs. What are some other possible benefits of such a group for cross-cultural researchers? What disadvantages might arise from use of such a group?

Web Exercises

1. The World Bank offers numerous resources that are useful for comparative research. Visit the World Bank website at www.worldbank.org. Scroll down to the "Data" heading and then click on the "Browse Data *by Country*" link and then select one country. Now, write a brief summary of changes reported in the indicators ranging from CO_2 emissions to total population. Then, compare these data with those for another country, and summarize the differences and similarities you have identified between the countries during recent decades.
2. The U.S. Bureau of Labor Statistics (BLS) website provides extensive economic indicator data for regions, states, and cities. Go to the BLS web page that offers statistics by location: http://stats.bls.gov/eag. Now, click on a region and explore the types of data that are available. Write out a description of the steps you would have to take to conduct a comparative analysis using the data available from the BLS website.
3. The U.S. Census Bureau's home page can be found at www.census.gov. This site contains extensive reporting of census data, including population data, economic indicators, and other information acquired through the U.S. Census. This website allows you to collect information on numerous subjects and topics. This information can then be used to make comparisons between different states or cities. Comparative analysis is facilitated by the "QuickFacts" option on the home page. Where it asks you to "Select a state to begin," choose your own state and copy down the percentages of the population by age, race and Hispanic origin, foreign born, education, and persons in poverty, as well as median household income. Repeat this process for one or two other states. Write a one-page report summarizing your findings about similarities and differences between these states and propose an explanation for at least some of the differences.

Video Interview Questions

1. What caused Cinzia Solari's research question to change? What was the comparative element in her research?
2. How did Solari build rapport between herself and the migrant workers she was trying to research? Why is this step important when doing qualitative research?

SPSS Exercises

1. In this exercise, you will use Mill's method of agreement to examine international differences in the role of labor unions. For this cross-sectional comparative analysis, you will use the ISSP data set on Work Orientations III 2005. Type in the following URL: http://www.jdsurvey.net/jds/jdsurveyAnalisis.jsp?ES_COL=127&Idioma=I&SeccionCol=06&ESID=453, which contains results of an international survey involving respondents in more than 25 countries.
    a. First, go to the Work Orientations III 2005 website and select three countries of interest. State a reason for choosing these three countries. State a hypothesis specifying which countries you expect to be relatively supportive of labor unions and which countries you expect to be relatively unsupportive of labor unions. Explain your reasoning.
    b. Once you have selected (checked) the countries you would like to compare, click on "Confirm selection" in the top left corner. Then, scroll down the list of questions and click on V25 (Without trade unions the working conditions of employees would be much worse than they are) or some other question of interest. Now click "NEXT" and then choose "Cross-tabs."
    c. Now review the crosstab of V25 (or whatever variable you selected) by country (country should be the column variable, and the table should show column percents). (It may also be useful to explore the "graphs" tab, which provides a visual comparison.) See if on this basis you can develop a tentative explanation of international variation in support for labor unions using Mill's method of agreement. You might also want to explore the question V24 ("Trade unions are very important for the job security of employees") to further compare international variation.
    d. Discuss the possible weaknesses in the type of explanation you have constructed, following John Stuart Mill. Propose a different approach for a comparative historical analysis.
2. How do the attitudes of immigrants to the United States compare with those of people born in the United States? Use the GSS2016 or GSS2016x or GSS2016x_reduced file and request the cross-tabulations (in percentage form) of POLVIEWS3, BIBLE, SPKATH by COHORTS (with COHORTS as the column variable). Inspect the output. Describe the similarities and differences you have found.
3. Because the GSS file is cross-sectional, we cannot use it to conduct historical research. However, we can develop some interesting historical questions by examining differences in the attitudes of Americans in different birth cohorts.
    a. Inspect the distributions of the same set of variables. Would you expect any of these attitudes and behaviors to have changed over the 20th century? State your expectations in the form of hypotheses.
    b. Request a cross-tabulation of these variables by birth COHORTS. What appear to be the differences between the cohorts? Which differences do you think result from historical change, and which do you think result from the aging process? Which attitudes and behaviors would you expect to still differentiate the baby boom generation and the post-Vietnam generation in 20 years?

Developing a Research Proposal

Add a historical or comparative dimension to your proposed study (Exhibit 3.10, #13 to #17).

1. Consider which of the four types of comparative or historical methods would be most suited to an investigation of your research question. Think of possibilities for qualitative and quantitative research on your topic with the method you prefer. Will you conduct a variable-oriented or case-oriented study? Write a brief statement justifying the approach you choose.
2. Review the possible sources of data for your comparative or historical project. Search the web and relevant government, historical, and international organization sites or publications. Search the social science literature for similar studies and read about the data sources that they used.

3. Specify the hypotheses you will test or the causal sequences you will investigate. Describe what your cases will be (nations, regions, years, etc.). Explain how you will select cases. List the sources of your measures, and describe the specific type of data you expect to obtain for each measure. Discuss how you will evaluate causal influences, and indicate whether you will take a nomothetic or idiographic approach to causality.
4. Review the list of potential problems in comparative and historical research and content analysis, and discuss those that you believe will be most troublesome in your proposed investigation. Explain your reasoning.

Chapter 16 Summarizing and Reporting Research

- Research That Matters, Questions That Count
- Writing Research
- Displaying Research
- Reporting Research
  - Journal Articles
- Research in the News: Do Preschool Teachers Need to Be College Graduates?
  - Applied Research Reports
    - *Findings From California's Welcome Baby Program*
    - *Limitations*
    - *Conclusions*
    - *Framing an Applied Report*
  - Research Posters
  - Reporting Quantitative and Qualitative Research
- Ethics, Politics, and Research Reports
- Careers and Research
  - Communicating With the Public
  - Plagiarism
- Conclusions

You learned in Chapter 2 that research is a circular process, so it is appropriate that we end this book where we began. The stage of reporting research results is also the point at which the need for new research is identified. It is the time when, so to speak, "the rubber hits the road"—when we have to make our research make sense to others. To whom will our research be addressed? How should we present our results to them? Should we seek to influence how our research report is used?

The primary goals of this chapter are to guide you in writing worthwhile reports of your own, displaying findings, and communicating with the public about research. This chapter gives particular attention to the writing process itself and points out how that process can differ when writing qualitative versus quantitative research reports. I will conclude by considering some of the ethical issues unique to the reporting process, with special attention to the problem of plagiarism.

Research That Matters, Questions That Count

Child maltreatment is a continuing problem in the United States, although its prevalence has decreased somewhat in the past three decades. Early childhood home visiting (ECHV) has become the most widely used type of program designed to prevent such maltreatment, but evidence of its effectiveness has been mixed. Steven Ondersma at Wayne State University and eight collaborators throughout the United States developed a new approach to ECHV that they hypothesized would be more effective. Their article in the journal *Child Maltreatment* reported the results of their experimental test of this new approach.

They called the new approach the e-Parenting Program (e-PP) because it added a computer component to a standard home visiting program for new mothers. At two sites within the Healthy Families Indiana program, new mothers identified as at-risk for child maltreatment who consented were assigned randomly to (1) the e-PP program (for a total of eight 20-minute sessions on a laptop) in addition to standard ECHV (weekly visits) for one year, (2) the standard ECHV program, or (3) a control condition that consisted of referral to community services. Outcome measures included an index of harsh parenting and measures of the following major risk factors for maltreatment: postnatal depression, substance abuse, interpersonal violence, and quality of home environment. These measures were collected at baseline, 6 months after the intervention began, and after the intervention's end at 12 months.

Ondersma and his colleagues did not find any effect of the enhanced program or the standard program on child maltreatment and only a few indications of reduction of risk factors due to participation in either e-PP or ECHV. They concluded that their computer-enhanced intervention program was not effective, but speculated that a software-based program with a more specific focus or one that was initiated more quickly after birth might have more impact.

1. Were you surprised by the study's finding of a lack of effect? What prior attitudes or experiences might have shaped your initial expectations? Can you set aside these expectations as you read the article and consider the strengths and weaknesses of the research design?
2. The article is organized in the standard major sections expected in a research article: Introduction (not stated), Method, Results, and Discussion. Try summarizing the major points in each section (and subsection). How easy is it to follow the writing in each section? What edits would you suggest to improve clarity?

In this chapter, you will learn how research reports and academic articles are organized and written. By the end of the chapter, you will know how to approach different tasks involved in reporting research and how to identify limitations in research reports. As you read the chapter, review the 2017 *Child Maltreatment* article by Steven Ondersma and his colleagues at the *Investigating the Social World* study site and complete the related interactive exercises for Chapter 16 at edge.sagepub.com/schutt9e.

Source: Ondersma, Steven J., Joanne Martin, Beverly Fortson, Daniel J. Whitaker, Shannon Self-Brown, Jessica Beatty, Amy Loree, David Bard, and Mark Chaffin. 2017. "Technology to Augment Early Home Visitation for Child Maltreatment Prevention: A Pragmatic Randomized Trial." *Child Maltreatment.* doi:10.1177/1077559517729890

Writing Research

The goal of research is not just to discover something but also to communicate that discovery to a larger audience: other social scientists, government officials, your teachers, the general public—perhaps several of these audiences. Whatever the study's particular outcome, if the intended audience for the research comprehends the results and learns from them, the research can be judged a success. If the intended audience does not learn about the study's results, the research should be judged a failure—no matter how expensive the research, how sophisticated its design, or how much you (or others) invested in it.

Successful research reporting requires both good writing and a proper publication outlet. We will first review guidelines for successful writing before we look at particular types of research publications.

Consider the following principles formulated by experienced writers (Booth, Colomb, and Williams 1995:150–151):

- Respect the complexity of the task and don't expect to write a polished draft in a linear fashion. Your thinking will develop as you write, causing you to reorganize and rewrite.
- Leave enough time for dead ends, restarts, revisions, and so on, and accept the fact that you will discard much of what you write.
- Write as fast as you comfortably can. Don't worry about spelling, grammar, and so on until you are polishing things up.
- Ask anyone whom you trust for reactions to what you have written.
- Write as you go along, so that you have notes and report segments drafted even before you focus on writing the report.

It is important to outline a report before writing it, but neither the report's organization nor the first written draft should be considered fixed. As you write, you will get new ideas about how to organize the report. Try them out. As you review the first draft, you will see many ways to improve your writing. Focus particularly on how to shorten and clarify your statements. Make sure that each paragraph concerns only one topic. Remember the golden rule of good writing: Writing is revising!

You can ease the burden of report writing in several ways:

- Draw on the research proposal and on project notes.
- Use a word processing program on a computer to facilitate reorganizing and editing.
- Seek criticism from friends, teachers, and other research consumers before turning in the final product.

I find it helpful at times to use what I call reverse outlining: After you have written a first complete draft, outline it on a paragraph-by-paragraph basis, ignoring the section headings you used. See if the paper you wrote fits the outline you planned. Consider how the organization could be improved.

Most important, leave yourself enough time so that you can revise, several times if possible, before turning in the final draft. Here are one student's reflections on writing and revising:

> I found the process of writing and revising my paper longer than I expected. I think it was something I was doing for the first time—working within a committee—that made the process not easy. The overall experience was very good, since I found that I have learned so much. My personal computer also did help greatly.
>
> Revision is essential until complete clarity is achieved. This took most of my time. Because I was so close to the subject matter, it became essential for me to ask for assistance in achieving clarity. My committee members, English editor, and fellow students were extremely helpful. Putting it on disk was also, without question, a timesaver. Time was the major problem.
>
> The process was long, hard, and time-consuming, but it was a great learning experience. I work full time so I learned how to budget my time. I still use my time productively and am very careful of not wasting it. (Graduate Program in Applied Sociology 1990:13)

Reverse outlining: Outlining the sections in an already written draft of a paper or report to improve its organization in the next draft.

For more suggestions about writing, see Howard Becker (2007), Booth et al. (2008), Lee Cuba (2002), William Strunk Jr. and E. B. White (2000), and Kate Turabian (2007).

Displaying Research

Chapter 9 introduced some conventions for reporting quantitative research results, including tabular displays and graphs. You can improve your displays of quantitative data in research reports by using the *Chart Builder* feature of SPSS (Aldrich and Rodriguez 2013).

Open the Chart Builder by clicking on "Graphs" and then "Chart Builder" in the SPSS Data Editor screen, with the GSS2016x data set. Exhibit 16.1 shows the SPSS Data Editor window as it appears at that point. To build a simple bar chart, click the "Gallery" option and then "Bar" in the "Choose from" list and drag the picture of the bar chart to the "Chart Preview" window. Now drag and drop the categorical variable whose distribution you would like to display (try "owngun") into the "X-Axis?" label box. If you would like to alter the graph or additional elements, click on the "Element Properties" tab and make desired changes (try changing the Bar 1 label to Percent and click "Apply" and then for the X-Axis 1 Bar, highlight and then exclude—with a red X—the category REFUSED and click "Apply") and then click "Close" and "OK." See Exhibit 16.2.

Chart Builder can also be used to create figures that display the association between variables. Exhibit 16.3 shows the association between political views and race using the 3-D Bar chart option with General Social Survey (GSS) 2016 data.

Graphic displays will help draw attention to important aspects of your findings and simplify interpretation of basic patterns. You should use graphic displays judiciously in a report, but use them you must if you want your findings to be accessible to the widest possible audience.

Presentations using PowerPoint or other presentation programs have become the norm for displaying research findings in the classroom, in professional meetings, and in public. When you display your research in this way, it's important to consider how to convey the key information without overloading or confusing your audience. SAGE author Vera Toepoel (2016:214) gives some straightforward guidelines:

Exhibit 16.1 Building a Chart With Chart Builder

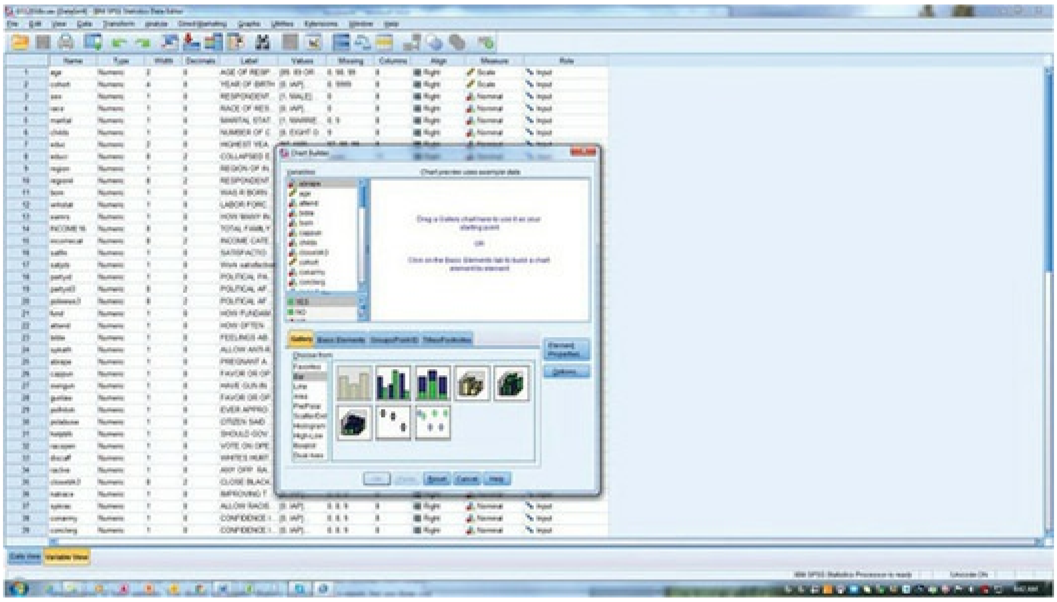


Exhibit 16.2 Adding Chart Features With Properties

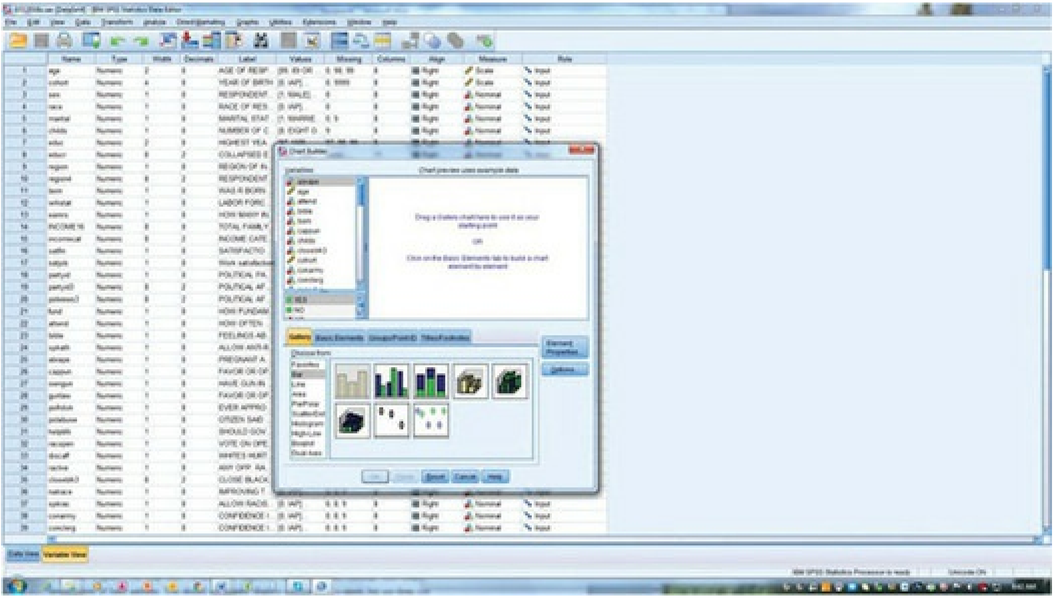
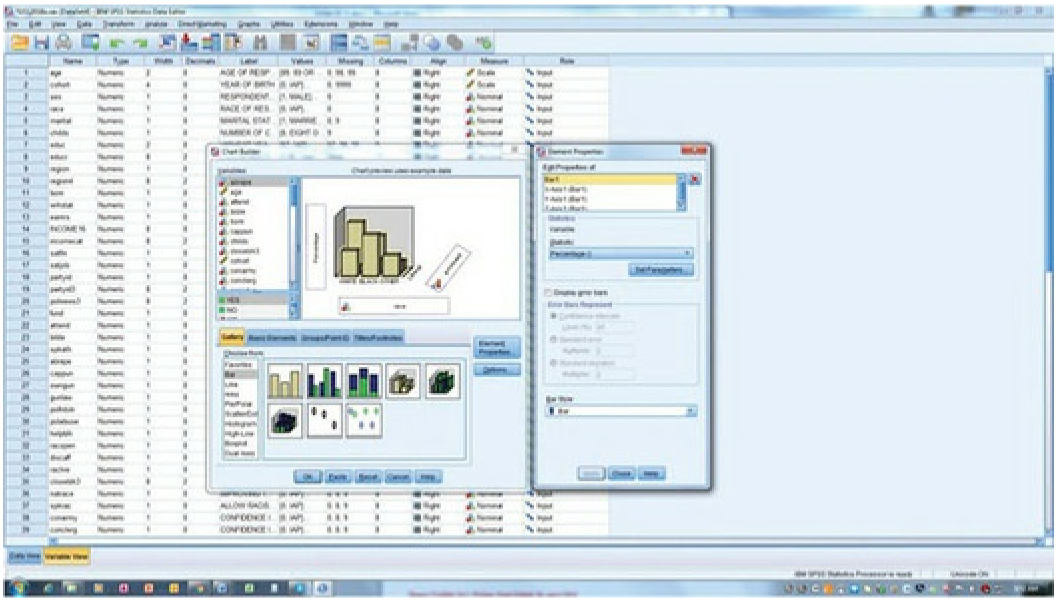


Exhibit 16.3 Displaying a Bivariate Association With Chart Builder



- Make sure that a slide contains only one idea, theory, or example.
- Do not spend too much time on each slide (1–2 minutes maximum)
- Give each slide a title that clearly explains what you are talking about.
- Do not use too many words on one slide (7 words × 7 rows maximum)
- Use bullet points for each separate point.
- Use cartoons, animation, fonts, and colors to increase the clarity of the point without impeding readability or looking messy.
- Cite sources and include references on the last slide.

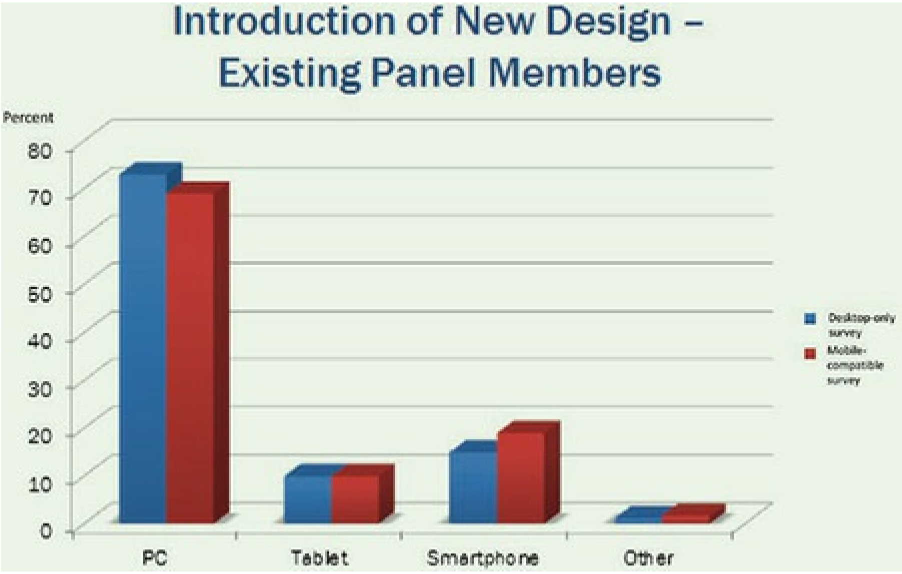
Exhibit 16.4 gives an example of one slide from one of Toepoel's own PowerPoint presentations with two colleagues. You can review that entire presentation to the European Survey Research Association in 2015 (and learn more about online surveys) at the book's study site (edge.sagepub.com/schutt9e).

Data visualization is a rapidly developing field that focuses on how to represent and present data to facilitate understanding (Kirk 2016:19). It goes a step further than just picking a chart type from a PowerPoint menu or statistical package like SPSS, by considering carefully how each element of a data presentation can be used to maximize understanding of what it contains. Understanding is most likely to be achieved in a data display that is

Data visualization: A rapidly developing field that focuses on how to represent and present data to facilitate understanding.

- *Trustworthy.* The basic pattern is not distorted to change the viewer's impression. For example, adding a three-dimensional appearance to figures in a chart can make them appear larger than they actually are.

Exhibit 16.4 Attractive Chart in PowerPoint



Source: Reprinted with permission from Vera Toepoel.

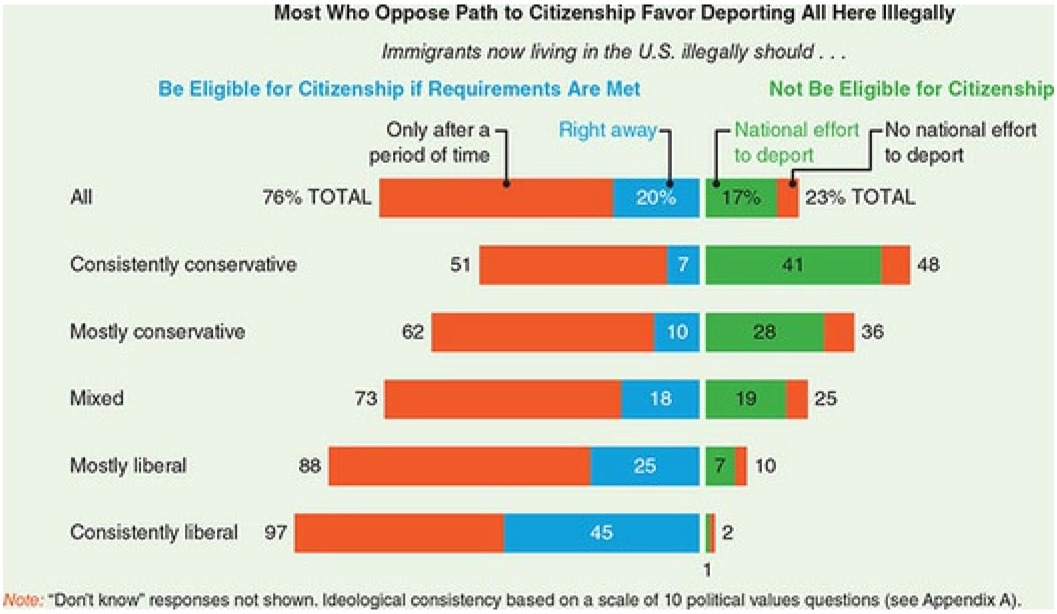
- *Accessible.* It should be easy for the viewer to see the key elements in the data and the relationships between them. Decisions about how simple a display needs to be have to take into account the expertise of the intended audience and the time they will have to view the display.
- *Elegant.* The design should attract and sustain attention. It's easy to overdo it, but try to make the design engaging as well as informative.

SAGE author Andy Kirk provides many examples in his book *Data Visualisation* and its website, http://book.visualisingdata.com/home. As one example, from a Pew Research Center study used by Kirk, Exhibit 16.5 shows how "back-to-back bar charts" can provide much more information than the simple bar charts you learned about in Chapter 9 while remaining easy to understand.

Reporting Research

You begin writing your research report when you are working on the research proposal and writing your literature review (see Chapter 2). You will find that the final report is much easier to write, and more adequate, if you write more material for it as you work out issues during the project. It is very disappointing to discover that something important was left out when it is too late to do anything about it. Students and professional researchers alike frequently leave final papers and reports until the last possible minute, often for understandable reasons, including other coursework and job or family responsibilities. But be forewarned: *Waiting until the last minute to write the report will not result in the best possible final product.*

Exhibit 16.5 Back-to-Back Bar Chart on Political Polarization



Source: Political Polarization in the American Public, Sedtion 4: Political Compromise and divisive Policy Debates.

Pew Research Center, Washington, DC. (June, 2014.) http://www.people-press.org/2014/06/12/section-4-political-compromise-and-divisive-policy-debates/

The organization of your research report will depend to some extent on the audience for which you are writing and the type of research you have conducted. Articles that will be submitted to an academic journal will differ from research reports written for a funding agency or for the general public. Research reports based on qualitative research will differ in some ways from those based on quantitative research. Students writing course papers are often required to structure their research reports using the journal article format, and they may be asked to present their results differently if they have used qualitative (or mixed) methods. The following sections outline the major issues to consider.

Journal Articles

Writing for academic journals is perhaps the toughest form of writing because articles are submitted to several experts in your field for careful review—anonymously, with most journals—before acceptance for publication. Perhaps it wouldn't be such an arduous process if so many academic journals did not have rejection rates in excess of 90% and turnaround times for reviews that are usually several months. Even the best articles, in the judgment of the reviewers, are most often given a "revise and resubmit" after the first review and then are evaluated all over again after the revisions are concluded.

In the News Research in the News: Do Preschool Teachers Need to Be College Graduates?

For Further Thought?

Considerable evidence indicates that high-quality early childhood programs can have long-term benefits for disadvantaged children, as reflected in research summarized in a report by Nobel Prize–winning University of Chicago economics professor James Heckman (https://heckmanequation.org/resource/research-summary-lifecycle-benefits-influential-early-childhood-program/). But does requiring preschool teachers to have a college degree result in higher program quality? It's quite an ongoing debate, since educational requirements also affect teacher pay and the availability of qualified teachers. Research indicates an association between teacher education and program quality, but "there has never been a large high-quality study, like a controlled trial that randomly placed children in a classroom with a college-educated teacher or not—and that also controlled for other variables that influence quality."

1. What type of study would you recommend to investigate the impact of preschool teachers' education on the quality of their teaching—and their effect on students? What would be some of the challenges to implementing your study design? Is it feasible?
2. In what ways could you report the results of research on the value of early childhood education programs in order to inform the public? Do you think that researchers should try to influence the public debate through their reports, or just stick to publishing in academic journals?

News source: Miller, Claire Cain. 2017. "The Perils of 'Bite Size' Science." *The New York Times,* April 7.

But journal article review procedures have some important benefits. Most important is the identification of areas in need of improvement, as the eyes of the author(s) are replaced by those of previously uninvolved subject-matter experts and methodologists. A good journal editor makes sure that he or she has a list of many different types of experts available for reviewing whatever types of articles the journal is likely to receive. There is a parallel benefit for the author(s): It is always beneficial to review criticisms of your own work by people who know the field well. It can be a painful and time-consuming process, but the entire field moves forward as researchers continually critique and suggest improvements in each other's research reports.

Are you familiar with the expression "no pain, no gain"? It applies here, too. Hundreds of online journals masquerade as scholarly journals but are simply money-making ventures for their “publishers” (Anderson 2017). These journals solicit articles from people eager to publish their papers, who then may pay as much as thousands of dollars to have their work included in an upcoming issue of the “open access” online journal. There is no peer review process, and often no real copyediting, so that articles can be published with bizarre claims and ungrammatical prose. Studies of journal practices involving submitting phony (and absurd) articles have found that such “predatory journals” will accept anything—for a fee (Bohannon 2013). They also recruit widely for editors, and research shows that they will appoint anyone to their editorial boards without regard to qualifications (Sorowski et al. 2017). The problem is not with open access journals per se, for there are some that are highly respected; the problem is with those that do not use a professional peer review process and are just seeking revenue.

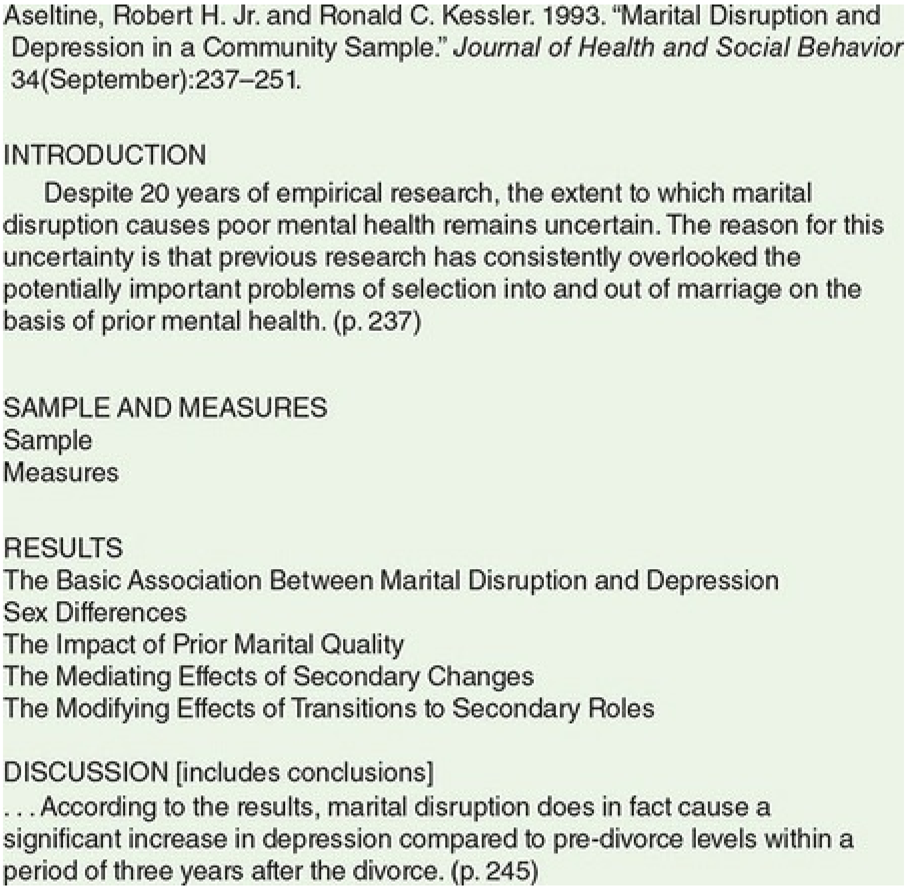
Exhibit 16.6 presents an outline of the sections in an academic journal article, with some illustrative quoted text. It is essential to begin with a clear abstract of the article, which summarizes in one paragraph the research question, prior research, study methods and major findings, and key conclusions. Many others who search the literature about the topic of your article will never read the entire article unless they are persuaded by the abstract that the article provides worthwhile and relevant information. The article’s introduction highlights the importance of the problem selected—in Exhibit 16.6 that’s the relationship between marital disruption (divorce) and depression. The introduction, which in this case includes the literature review, also identifies clearly the gap in the research literature that the article is meant to fill: the untested possibility that depression might cause marital disruption rather than, or in addition to, marital disruption causing depression. Literature reviews in journal articles should be integrated reviews that highlight the major relevant findings and identify key methodological lessons from the prior research as a whole, rather than presenting separate summaries of prior research studies (see Chapter 2). The findings section in Exhibit 16.6 (titled “Results”) begins by presenting the basic association between marital disruption and depression. Then it elaborates on this association by examining sex differences, the impact of prior marital quality, and various mediating and modifying effects. Tables and perhaps graphs are used to present the data corresponding to each of the major findings in an easily accessible format. As indicated in the combined Discussion and Conclusions section in the exhibit, the analysis shows that marital disruption does indeed increase depression and specifies the time frame (3 years) during which this effect occurs.

These basic article sections present research results well, but many research articles include subsections tailored to the issues and stages in the specific study being reported. Most journals require a short abstract at the beginning, which summarizes the research question and findings. Most research articles include a general Methodology section that will include subsections on measurement and sampling. A Conclusions section is often used to present the most general conclusions, reflections, and limitations, but some precede that with a general Discussion section.

Applied Research Reports

Applied research reports are written for a different audience from the professional social scientists and students who read academic journals. Typically, an applied report is written with a wide audience of potential users in mind and to serve multiple purposes. Often, both the audience and the purpose are established by the agency or other organization that funded the research project on which the report is based. Sometimes, the researcher may use the report to provide a broad descriptive overview of the study findings, which will be presented more succinctly in a subsequent journal article. In either case, an applied report typically provides much more information about a research project than does a journal article and relies primarily on descriptive statistics rather than only those statistics useful for the specific hypothesis tests that are likely to be the primary focus of a journal article.

Exhibit 16.6 Sections in a Journal Article



Aseltine, Robert H. Jr. and Ronald C. Kessler. 1993. "Marital Disruption and Depression in a Community Sample." *Journal of Health and Social Behavior* 34(September):237–251.

INTRODUCTION

Despite 20 years of empirical research, the extent to which marital disruption causes poor mental health remains uncertain. The reason for this uncertainty is that previous research has consistently overlooked the potentially important problems of selection into and out of marriage on the basis of prior mental health. (p. 237)

SAMPLE AND MEASURES

Sample

Measures

RESULTS

The Basic Association Between Marital Disruption and Depression

Sex Differences

The Impact of Prior Marital Quality

The Mediating Effects of Secondary Changes

The Modifying Effects of Transitions to Secondary Roles

DISCUSSION [includes conclusions]

. . . According to the results, marital disruption does in fact cause a significant increase in depression compared to pre-divorce levels within a period of three years after the divorce. (p. 245)

Source: Aseltine, Robert H., Jr. and Ronald C. Kessler, 1993. "Marital Disruption and Depression in a Community Sample." Journal of Health and Social Behavior 34(September):237–251.

One of the major differences between an applied research report and a journal article is that a journal article must focus on answering a particular research question, whereas an applied report is likely to have the broader purpose of describing a wide range of study findings and attempting to meet the diverse needs of multiple audiences for the research. But a research report that simply describes "findings" without some larger purpose in mind is unlikely to be effective in reaching any audience. Anticipating the needs of the intended audience(s) and identifying the ways in which the report can be useful to them will result in a product that is less likely to be ignored.

Findings From California's Welcome Baby Program

A good example of applied research reporting comes from the Urban Institute's report on an evaluation of a California program for new mothers called the Welcome Baby program (Sandstrom et al. 2015). With funding from a nonprofit child-advocacy organization, First 5 LA, Heather Sandstrom and her colleagues at the Urban Institute, and Todd Franke and others at the University of California, Los Angeles, conducted a mixed-methods evaluation of the implementation and impacts of the Welcome Baby program in metropolitan Los Angeles (Metro LA) from 2010 to 2015.

Exhibit 16.7 outlines the sections in the applied research report prepared by the Urban Institute to present their findings from the evaluation, with some illustrative text and one exhibit. The report is described as presenting results from the final wave of data collection in the Welcome Baby study—three years postpartum—and from longitudinal analyses of change over time. Like most applied reports, it begins with an executive summary that presents the highlights from each section of the report, including major findings, in a brief format. The body of the report summarizes prior research on home visiting; describes the intervention, the research methods used, and participant characteristics; and then presents findings about program impact. It concludes with Limitations and Discussion sections and includes several appendices with more study details.

Their 36-month study sought "to examine whether and how Welcome Baby in Metro LA improves the health, development, and well-being of very young children and their families" (Sandstrom et al. 2015:ii). The report describes the components of the surveys conducted at 12, 24, and 36 months postpartum (Sandstrom et al. 2015:ii):

1. A 90-minute parent interview that draws upon several validated scales designed to measure key aspects of parental well-being, the home environment, and children's health and development
2. A 10-minute observational assessment of a semi-structured, parent-child play session designed to measure the quality of parent-child interactions
3. A home observation checklist that assesses the quality of the home environment and immediate neighborhood
4. A direct assessment of child height and weight, at 24 and 36 months only

Together the measures examine seven key outcome domains: quality of the home environment, parenting and the parent-child relationship, child development, child nutrition, maternal and child health care and coverage, maternal mental health, and family well-being. Distinct research questions are posed for the longitudinal analysis of change over time (Sandstrom et al. 2015:18):

1. What is the impact of Welcome Baby on child and family outcomes over time? Do significant benefits among the intervention group grow or diminish as children age?
2. Is Welcome Baby associated with greater benefits for certain subgroups of children and families? Do the outcomes of families with first-time, partnered, and better-educated mothers differ significantly from those of other families?
3. How does the timing of enrollment in Welcome Baby affect child and family outcomes? Does initiating Welcome Baby prenatally versus postpartum predict better child and family outcomes when children are 12 through 36 months old?
4. How does the dosage of Welcome Baby services affect child and family outcomes? Do families who complete the 9-month home visit demonstrate better child and family outcomes when children are 12 through 36 months old than families who do not complete this visit?

The report summarizes findings from the data collected prior to 36 months:

> Outcomes identified from previous waves of the Child and Family Survey have been consistent with expected effects of the home visiting intervention. For instance, findings show Welcome Baby participants have higher rates of attempted breastfeeding and exclusive breastfeeding, providing evidence that parent coaches, hospital liaisons and nurses are effectively supporting mothers in their efforts to breastfeed their babies. (Sandstrom et al. 2015:6)

Exhibit 16.7 Sections in an Applied Report

Sandstrom, Heather, Sarah Benatar, Erica Greenberg, Ian Hill, Embry Howell, Timothy Triplett, Margaret Wilkinson, Todd Franke, Christina Christie, Maria Lourdes Brown, and Jennifer Ho. 2015. Welcome Baby Home Visiting: Findings from the 36-Month Child & Family Survey and 3-Year Longitudinal Results. Urban Institute. Retrieved from http://www.first5la.org/postfiles/files/36-Month%20Outcomes%20Report.pdf

EXECUTIVE SUMMARY

Welcome Baby is a free and voluntary home visiting program recently established at fourteen hospitals in communities throughout Los Angeles County.

Overview of the Evaluation of Welcome Baby in Metro LA

Key Findings

Table 1. Significant Outcomes at 36 Months Associated with Welcome Baby

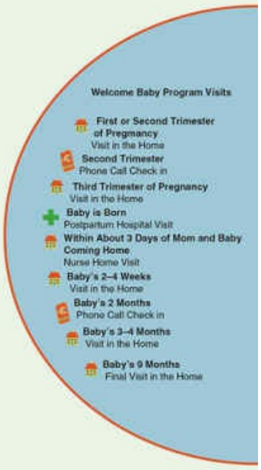
Parent Outcomes	Child Outcomes
Stronger maternal responsiveness	Greater communication skills
Stronger maternal encouragement	Greater social competence
Stronger maternal affection	Greater engagement and attention
Stronger maternal teaching	Greater personal-social skills
Lower parenting stress	Lower body mass index
Higher rates of dental insurance	

Study Limitations

Implications for the Welcome Baby Program

I. INTRODUCTION

This report presents findings from the third and final wave of data collection—three years postpartum—as well as a discussion of longitudinal analyses that pool data across all three waves to estimate variation in effects over time. The...



II. OVERVIEW OF WELCOME BABY

Welcome Baby home visiting focuses on the provision of education and support for pregnant women and mothers of newborns.

III. PREVIOUS EVALUATION FINDINGS

IV. METHODS

A. Sample Recruitment and Retention

B. Description of 36-Month Child and Family Survey

The survey is composed of four parts: 1) a parent interview that draws on several validated scales designed to measure key aspects of parental well-being, the home environment, and children's health and development; 2) a 10-minute observational assessment of a semi-structured, parent-child play session designed to measure the quality of parent-child interactions; 3) a home observation checklist that assesses the quality of the home environment and immediate neighborhood; and 4) a direct assessment of child height and weight using a stadiometer and scale brought to the home.

C. Data Collection Procedures for 36-Month Child and Family Survey

D. Research Questions and Analytic Approach

V. STUDY FINDINGS

A. Cross-Sectional Analysis of the 36-Month Child & Family Survey

Quality of Parenting and Parent Child Interactions

Overall, Welcome Baby participants display more positive parenting behaviors when interacting with their three-year-olds than do comparison group participants, and likewise, their children display more positive behaviors towards them.

Child Development
Home Environment and Activities
Child Nutrition and BMI
Maternal and Child Physical Health and Health Care Coverage
Maternal Mental Health
Family Well-Being and Sources of Economic Support

Similar to previous survey findings, Welcome Baby clients do not significantly vary from comparison group participants in terms of material hardship, food insecurity, and the use of public benefits at 36 months.

B. Timing and Dosage Effects: A Sensitivity Analysis Among Mothers Enrolled in Welcome Baby Home Visiting

C. Longitudinal Analysis of Effects of Welcome Baby

Welcome Baby is associated with significant improvements in child communication skills and parental engagement in home learning activities at 12 months.

D. Timing and Dosage Effects: A Sensitivity Analysis of Longitudinal Findings

VI. STUDY LIMITATIONS

VII. DISCUSSION AND IMPLICATIONS

Despite being a relatively low-intensity home visiting program, administered to a diverse set of families, with outcomes being measured for more than two years beyond the completion of the program, findings have demonstrated that Welcome Baby contributes to:

- Positive parenting practices, such as engagement in home learning activities and maternal responsiveness, encouragement, and teaching behaviors towards children;
- Children's communication skills and social competence; and
- Select maternal and child mental and physical health markers, such as parenting stress and child BMI.

Additionally, findings show that women who are potentially at greater risk of negative outcomes—mothers with less than a high school education and single mothers—may be benefiting most from the program in terms of their mental health and parenting skills.

Appendix A: Key Measures from 36-Month Child and Family Survey
Appendix B: Measures Available for Longitudinal Analysis
Appendix C: Description of Key Measures from Child and Family Survey
Appendix D: Longitudinal Data File Creation
Appendix E: 36-Month Child and Family Survey
Appendix F: Heterogeneity in the Effects of Welcome Baby by Mothers' Characteristics

Source: Vernez et al. (1988).

The Findings section presents detailed outcomes at 36 months and for the longitudinal analysis of change over time.

> Overall, Welcome Baby participants display more positive parenting behaviors when interacting with their three-year-olds than do comparison group participants, and likewise, their children display more positive behaviors towards them. (Sandstrom et al. 2015:25)
>
> Welcome Baby participation is positively associated with children's communication skills and social competence at 36 months. (Sandstrom et al. 2015:32)
>
> Welcome Baby is associated with lower parental stress and greater perceived social support at 36 months, but consistent with previous survey findings, the program has no measurable effects on maternal depression. (Sandstrom et al. 2015:39)

Limitations

The researchers noted several limitations (Sandstrom et al. 2015:61–62):

1. This is a quasi-experimental design, with a comparison group of women residing in the community who were not offered the home visiting intervention. Participants were not randomly assigned, and the intervention and comparison groups differ on several demographic and socioeconomic characteristics. . . . [U]nmeasured differences could not be controlled for.
2. The comparison group sample size is smaller than initially targeted, and there was a low response-rate for this group. . . . [A] larger sample size might better detect slight group differences.
3. The evaluation follows children after they complete the Welcome Baby program to measure outcomes at 12-, 24-, and 36-months postpartum; however, no baseline or pre-intervention survey data were collected. . . . Without baseline data on the same outcome measures, the evaluation cannot make robust causal inferences about the effects of the intervention but rather estimates associations between program participation and later child and family outcomes.
4. The findings are not generalizable to all Welcome Baby clients given the study inclusion criteria. Study participants were required to be at least 18 years old at the

time of the 12-month survey, so young teenage mothers are not represented in the sample. . . . Further, this study was focused on Welcome Baby as implemented in Metro LA only.

5. The evidence suggests that program effects vary by timing of enrollment and program completion; however, these findings are limited by the fact that women self-selected to participate in the program prenatally and to complete the final visit of the program.
6. The large number of analyses conducted to test these associations increased the possibility of identifying a significant effect. However, it is possible that some estimated effects were found to be significant due to chance and were not the result of Welcome Baby participation.

Conclusions

The researchers concluded,

> Overall, these findings indicate that Welcome Baby is having important positive effects on women who are enrolled in the home visiting program, with indications that earlier and longer exposure to the intervention is associated with more significant effects. Additional analyses at 36 months have confirmed many earlier findings, and revealed a few new ones. Furthermore, longitudinal analyses that garner the strength of pooled samples add an additional layer of confidence to the results presented. In addition, as case study and focus group findings corroborate, overall satisfaction with the program is very high, and participants find value in the guidance and support they are receiving. . . . Since these positive findings are for a program that was evaluated during its infancy and early years of implementation, future cohorts arguably may enjoy even more positive results as Welcome Baby matures and becomes more systematic and consistent.
> (Sandstrom et al. 2015:66)

The Urban Institute report also emphasized the importance of the lessons learned from this study for new research to be conducted after the Welcome Baby program was expanded into 13 additional communities in Los Angeles County:

> This evaluation of a cohort of families who received Welcome Baby home visiting in the Metro LA pilot community has already provided important lessons for the expansion of Welcome Baby into 13 additional communities across Los Angeles County. The prospect of more rigorous study of the expanded intervention may lend additional credence to this study’s findings regarding the benefits that this program can achieve. In the interim, these longitudinal quasi-

> experimental findings demonstrate the promise that Welcome Baby home visiting offers to expectant and newly postpartum parents and their children. (Sandstrom et al. 2015:66)

Framing an Applied Report

What can be termed the front matter and the back matter of an applied report also are important. Applied reports usually begin with an executive summary: a summary list of the study's main findings, often in bullet fashion. Appendixes, the back matter, may present tables containing supporting data that were not discussed in the body of the report. Applied research reports also often append a copy of the research instrument(s).

An important principle for the researcher writing for a nonacademic audience is that the findings and conclusions should be engaging and clear. You can see how I did this in a report from a class research project I designed with my graduate methods students (and in collaboration with several faculty knowledgeable about substance abuse) (see Exhibit 16.8). These report excerpts indicate how I summarized key findings in an executive summary (Schutt et al. 1996:iv), emphasized the importance of the research in the introduction (Schutt et al. 1996:1), used formatting and graphing to draw attention to particular findings in the body of the text (Schutt et al. 1996:5), and tailored recommendations to our own university context (Schutt et al. 1996:26).

Front matter: The section of an applied research report that includes an executive summary, an abstract, and a table of contents.

Back matter: The section of an applied research report that may include appendixes, tables, figures, and the research instrument(s).

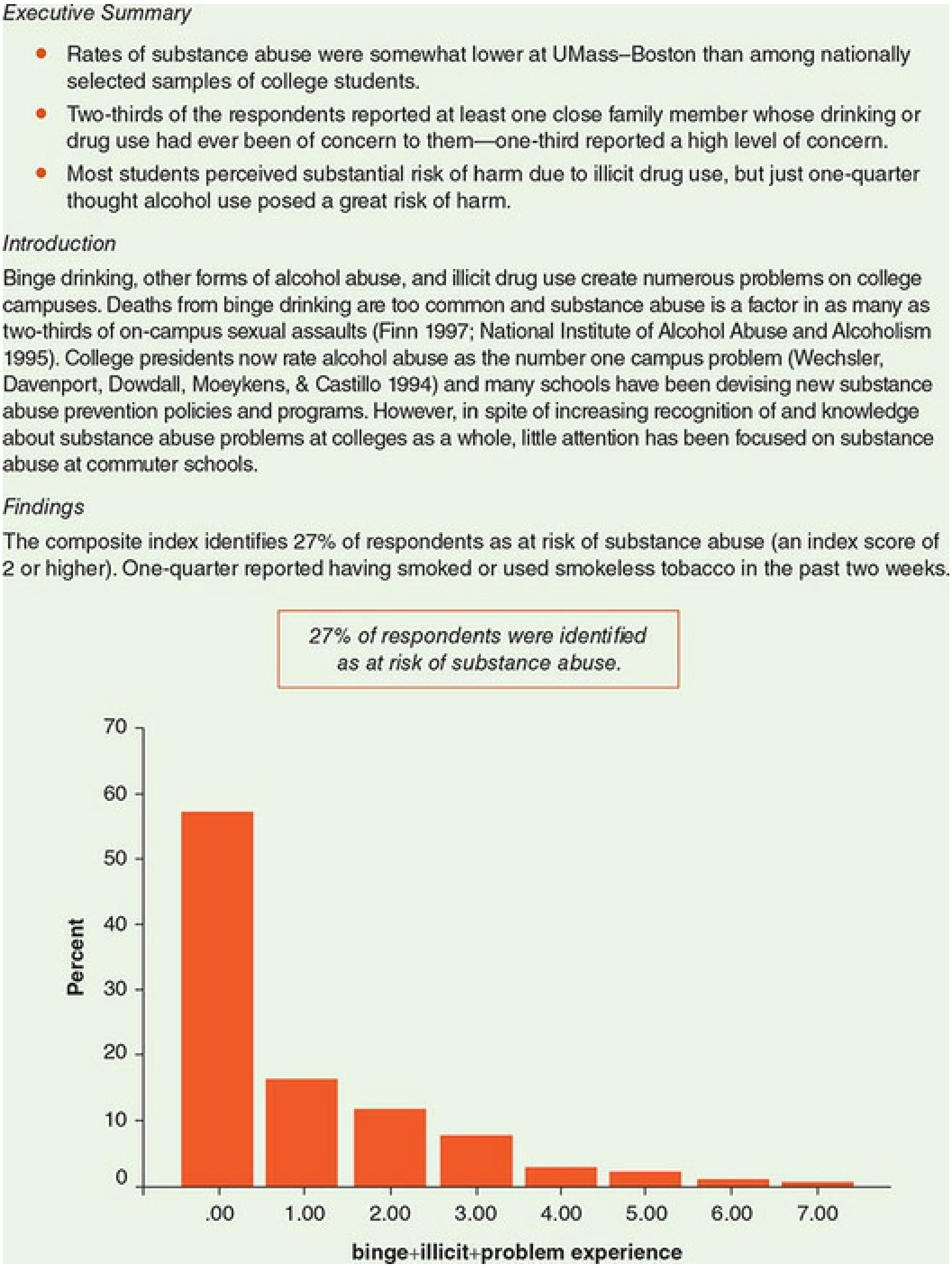
Research Posters

Research posters have long been a popular tool for summarizing research projects and findings at professional meetings in medicine and psychology, and professional associations in sociology now often include one or more poster sessions. A well-designed poster can provide much of the key information that would otherwise appear in a paper or presentation, while the poster session format can allow meeting participants to question presenters at length about their work and perhaps exchange information about their own, or just consider the findings from a distance.

Once again, SAGE author Vera Toepoel (2016:217) provides helpful guidelines:

- Completely fill the poster, avoiding clutter or overwhelming appearance.
- Include no more than 15 lines per page.
- Use bullets, but brief, on one point only, and not too many in one list.
- Use a font large enough for reading 3–6 feet away.

Exhibit 16.8 Student Substance Abuse, Report Excerpts



Executive Summary

- Rates of substance abuse were somewhat lower at UMass–Boston than among nationally selected samples of college students.
- Two-thirds of the respondents reported at least one close family member whose drinking or drug use had ever been of concern to them—one-third reported a high level of concern.
- Most students perceived substantial risk of harm due to illicit drug use, but just one-quarter thought alcohol use posed a great risk of harm.

Introduction

Binge drinking, other forms of alcohol abuse, and illicit drug use create numerous problems on college campuses. Deaths from binge drinking are too common and substance abuse is a factor in as many as two-thirds of on-campus sexual assaults (Finn 1997; National Institute of Alcohol Abuse and Alcoholism 1995). College presidents now rate alcohol abuse as the number one campus problem (Wechsler, Davenport, Dowdall, Moeykens, & Castillo 1994) and many schools have been devising new substance abuse prevention policies and programs. However, in spite of increasing recognition of and knowledge about substance abuse problems at colleges as a whole, little attention has been focused on substance abuse at commuter schools.

Findings

The composite index identifies 27% of respondents as at risk of substance abuse (an index score of 2 or higher). One-quarter reported having smoked or used smokeless tobacco in the past two weeks.

27% of respondents were identified as at risk of substance abuse.

Source: Schutt et al. (1996).

- Graphs and visuals usually are better than plain text.
- Use software designed to create posters.
- Ensure a customary flow, using top left to bottom right and numbered sections.
- Place most important or eye-catching information (usually results) in the center.

The University of Texas provides guidelines on their website (https://ugs.utexas.edu/our/poster/samples) and a number of good examples. Adam Pittman's poster on immigration and crime is a good model (see Exhibit 16.9). You can create a poster like this just using PowerPoint (with a large printer), but it's much easier if you use software designed for this purpose.

Reporting Quantitative and Qualitative Research

The requirements for good research reports are similar in many respects for quantitative and qualitative research projects. Every research report should include good writing, a clear statement of the research question, an integrated literature review, and presentation of key findings with related discussion, conclusions, and limitations. The outline used in Robert Aseltine and Ronald Kessler's (1993) report of a quantitative project may also be used by some authors of qualitative research reports. The Robert Wood Johnson research report also provides an example of how a research report of a mixed-methods study can integrate the results of analyses of both types of data. However, the differences between qualitative and quantitative research approaches mean that it is often desirable for research reports based on qualitative research to diverge in some respects from those reflecting quantitative research.

Reports based on qualitative research should be enriched in each section with elements that reflect the more holistic and reflexive approach of qualitative projects. The introduction should include background about the development of the researcher's interest in the topic, whereas the literature review should include some attention to the types of particular qualitative methods used in prior research. The methodology section should describe how the researcher gained access to the setting or individuals studied and the approach used to managing relations with research participants. The presentation of findings in qualitative studies may be organized into sections reflecting different themes identified in interviews or observational sessions. Quotes from participants or from observational notes should be selected to illustrate these themes, although qualitative research reports differ in the extent to which the researcher presents findings in summary form or uses direct quotes to identify key issues. The findings sections in a qualitative report may alternate between presentations of quotes or observations about the research participants, the researcher's interpretations of this material, and some commentary on how the researcher reacted in the setting, although some qualitative researchers will limit their discussion of their reactions to the discussion section.

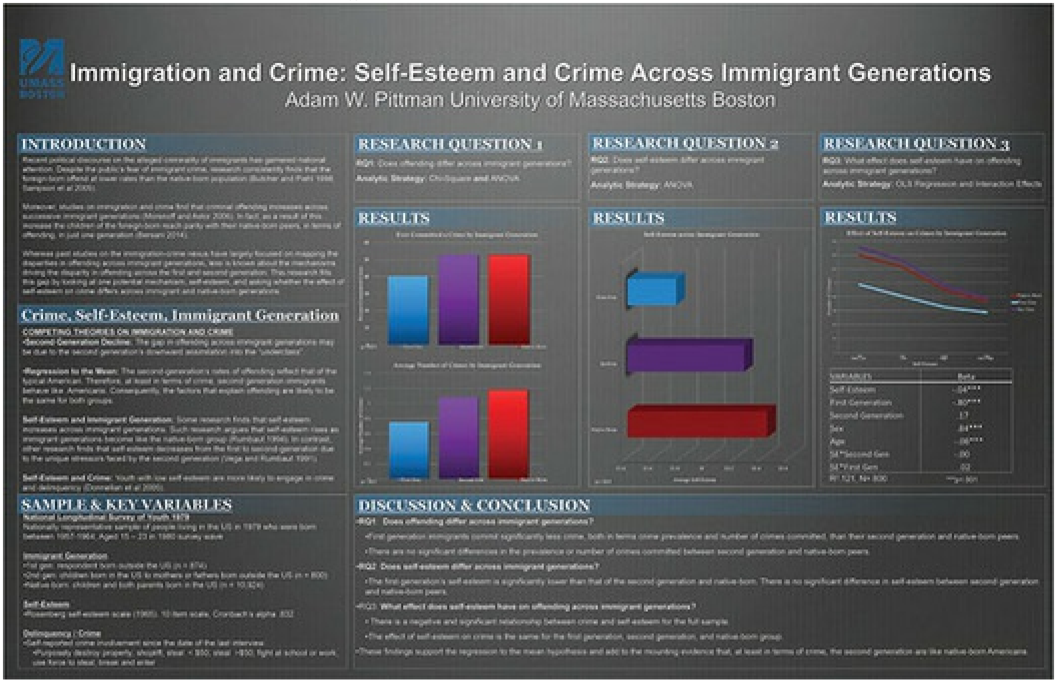
Reports on mixed-methods projects should include subsections in the methods section that introduce each method, and then distinguish findings from qualitative and quantitative analyses in the findings section. Some mixed-methods research reports may present analyses that use both qualitative and quantitative data in yet another subsection, but others may just discuss implications of analyses of each type for the overall conclusions in the discussions and conclusions sections (Dahlberg, Wittink, and Gallo 2010:785–791). When findings based on each method are presented, it is important to consider explicitly both the ways in which the specific methods influenced findings obtained with those methods and to discuss the implications of findings obtained using both methods for the overall study conclusions.

Ethics, Politics, and Research Reports

The researcher's ethical duty to be honest and open becomes paramount when reporting research results. Here are some guidelines:

- *Provide an honest accounting of how the research was carried out and where the initial research design had to be changed.* Readers do not have to know about every change you made in your plans and each new idea you had, but they should be informed about major changes in hypotheses or research design. If important hypotheses were not supported, acknowledge this, rather than conveniently forgetting to mention them (Brown and Hedges 2009:383). If a different approach to collecting or analyzing the data could have led to different conclusions, this should be acknowledged in the limitations section (Bergman 2008:588–590).

Exhibit 16.9 Poster Example



Immigration and Crime: Self-Esteem and Crime Across Immigrant Generations

Adam W. Pittman University of Massachusetts Boston

INTRODUCTION

RESEARCH QUESTION 1

RQ1: Does offending differ across immigrant generations?

Analytic Strategy: Chi-Square and ANOVA

RESEARCH QUESTION 2

RQ2: Does self-esteem differ across immigrant generations?

Analytic Strategy: ANOVA

RESEARCH QUESTION 3

RQ3: What effect does self-esteem have on offending across immigrant generations?

Analytic Strategy: OLS Regression and Interaction Effects

RESULTS

Crime, Self-Esteem, Immigrant Generation

SAMPLE & KEY VARIABLES

DISCUSSION & CONCLUSION

Source: Adam W. Pittman, University of Massachusetts, Boston.

- *Evaluate honestly the strengths and weaknesses of your research design.* Systematic evaluations suggest that the stronger the research design from the standpoint of establishing internal (causal) validity, the weaker the empirical support that is likely to be found for hypothesized effects (compare Weisburd, Lum, and Petrosino 2001:64). Finding support for a hypothesis tested with a randomized experimental design is stronger evidence than support based on correlations between variables measured in a cross-sectional survey.

- *Refer to prior research and interpret your findings within the body of literature resulting from that prior research.* Your results are likely to be only the latest research conducted to investigate a research question that others have studied. It borders on unethical practice to present your findings as if they are the only empirical information with which to answer your research question, yet many researchers commit this fundamental mistake (Bergman 2008:599). For example, a systematic evaluation of citation frequency in articles reporting clinical trial results in medical journals found that, on average, just 21% of the available prior research was cited (for trials with at least three prior articles that could have been cited) (Robinson and Goodman 2011:53). The result of such omission is that readers may have no idea whether your own research supports a larger body of evidence or differs from it—and so should be subject to even greater scrutiny.
- *Maintain a full record of the research project so that questions can be answered if they arise.* Many details will have to be omitted from all but the most comprehensive reports, but these omissions should not make it impossible to track down answers to specific questions about research procedures that may arise during data analysis or presentation. Tests of relationships that were conducted but not included in the report should be acknowledged.
- *Avoid "lying with statistics" or using graphs to mislead.* (See Chapter 9 for more on this topic.) There is a more subtle problem to be avoided, which is "cherry-picking" results to present. Although some studies are designed to test only one hypothesis involving variables that each are measured in only one way, many studies collect data that can be used to test many hypotheses, often with alternative measures. If many possible relationships have been examined with the data collected and only those found to yield a statistically significant result are reported, the odds of capitalizing on chance findings are multiplied. This is a major temptation in research practice and has the unfortunate result that most published findings are not replicated or do not stand up to repeated tests over time (Lehrer 2010:57). Every statistical test presented can be adequately understood only in light of the entire body of statistical analyses that led to that particular result.
- *Acknowledge the sponsors of the research.* This is important partly so that others can consider whether this sponsorship may have tempted you to bias your results in some way. Whether you conducted your research for a sponsor, or together with members of an underserved community, give research participants an opportunity to comment on your main findings before you release them to the public. Consider revising your report based on their suggestions or, if you disagree with their suggestions, include their comments in footnotes at relevant points in your report or as an appendix to it (Bledsoe and Hopson 2009:392).
- *Thank staff who made major contributions.* This is an ethical as well as a political necessity. Let's maintain our social relations!
- *Be sure that the order of authorship for coauthored reports is discussed in advance and reflects agreed-on principles.* Be sensitive to coauthors' needs and concerns.

Ethical research reporting should not mean ineffective reporting. You need to tell a coherent story in the report and avoid losing track of the story in a thicket of minuscule details. You do not need to report every twist and turn in the conceptualization of the research problem or the conduct of the research. But be suspicious of reports that don't seem to admit to the possibility of any room for improvement. Social science is an ongoing enterprise in which one research report makes its most valuable contribution by laying the groundwork for another, more sophisticated, research project. Highlight important findings in the research report but also use the research report to point out what are likely to be the most productive directions for future researchers.

Careers and Research

Julia Schutt, MA

Julia Schutt's first job after college was as the database manager for a large longitudinal study on schizophrenia onset at one local hospital site. As a psychology major, Schutt had favored courses in qualitative methods and was surprised by how much she enjoyed quantitative research. Eventually, however, she left that job to pursue a master's degree in international human rights, where she used a variety of qualitative approaches to study justice and systemic violence in Colombia.

Schutt spent a year working in human rights in Colombia after receiving her master's. She then decided to return to her hometown and focus on asking similar questions at a local level. She was hired as a policy associate at a small nonprofit dedicated to finding systemic solutions to social justice issues. At this organization, Schutt led a research project with the trial court system, where she was able to draw on her varied experience in quantitative and qualitative methods to help the court assess the needs of court users, especially those without lawyers, and to design an online self-help center based on those needs.

Within her first year, Schutt submitted a winning proposal to complete an annual evaluation of an online document assembly platform used by courts, legal aid organizations, self-represented litigants, and attorneys. She worked with the platform's developers to identify the organization's goals and project objectives to ensure that the evaluation assessed the program not only accurately but also in accordance with LHI's own priorities.

Her advice for current students is to "be a good listener when it comes to working with clients, consumers, and communities." She would encourage students to "embrace creativity and not be afraid to suggest trying something new to stakeholders. Think not only about how to keep clients happy, but also about how to challenge them. Positive changes typically require breaking away from the mold—be the one to show a new direction!"

Ethical research reporting should mean reporting findings regardless of whether they support your hypotheses, but that is often not the case. When you "don't find anything" or your hypotheses are not supported, this can be important information for other researchers, but it will be very hard—even impossible—to have an article about the "null findings" accepted by a journal. Journals (and authors) like to publish articles that show something new about social phenomena (Gough et al. 2017:99). There can also be an assumption that poor decisions in the research methods used are likely to have led to the lack of anticipated effects. Over time, this tendency means that the peer-reviewed journal literature can have a publication bias toward research that shows hypothesized effects, rather than showing that effects are not found. Since hypotheses are often not supported, researchers joke about wishing there was a *Journal of Null Findings* in which they could publish these undesired outcomes. Concerns about publication bias in medical research led the U.S. Congress in 1997 to pass a law requiring registration of clinical trials (health-related experiments) before enrolling research subjects (U.S. National Library of Medicine n.d.). Many medically oriented journals now require that all clinical trials be registered in the publicly available database maintained by the U.S. National Library of Medicine at clinicaltrials.gov. Authors must confirm at the time they submit an article to one of these journals that their research project was registered in the database.

Publication bias: The greater likelihood of publication of articles reporting tests of hypotheses that were supported.

Communicating With the Public

Even following appropriate guidelines such as these, however, will not prevent controversy and conflict over research on sensitive issues. The sociologist Peter Rossi (1999) recounts the controversy that arose when he released a summary of findings conducted in his 1989 study of homeless persons in Chicago (see Chapter 5). Despite important findings about the causes and effects of homelessness, media attention focused on Rossi's markedly smaller estimate of the numbers of homeless persons in Chicago compared with the "guesstimates" that had been publicized by local advocacy groups. "Moral of the story: Controversy is news, to which careful empirical findings cannot be compared" (Rossi 1999:2).

Does this mean that ethical researchers should avoid political controversy by sidestepping media outlets for their work? Many social scientists argue that the media offers one of the best ways to communicate the practical application of sociological knowledge and that when we avoid these opportunities, "some of the best sociological insights never reach policy makers because sociologists seldom take advantage of useful mechanisms to get their ideas out" (Wilson 1998:435).

The sociologist William Julius Wilson (1998:438) urges the following principles for engaging the public through the media:

1. Focus on issues of national concern, issues that are high on the public agenda.
2. Develop creative and thoughtful arguments that are clearly presented and devoid of technical language.
3. Present the big picture whereby the arguments are organized and presented so that the readers can see how the various parts are interrelated.

Ultimately each researcher must make a decision about the most appropriate and important outlets for his or her work.

Plagiarism

It may seem depressing to end a book on research methods with a section on plagiarism, but it would be irresponsible to avoid the topic. Of course, you may have a course syllabus detailing instructor or university policies about plagiarism and specifying the penalties for violating that policy, so I'm not simply going to repeat that kind of warning. You probably realize that the practice of selling term papers is revoltingly widespread (my search of "term papers for sale" on Google returned 901,000 websites on October 24, 2017)—for at least the first 10 pages of "hits," every site provided an opportunity to buy a paper, not a criticism of the practice. Instead of repeating the warnings, I will use this section to review the concept of plagiarism and to show how that problem connects to the larger issue of the integrity of social research. When you understand the dimensions of the problem and the way it affects research, you should be better able to detect plagiarism in other work and avoid it in your own.

You learned in Chapter 3 that maintaining professional integrity—honesty and openness in research procedures and results—is the foundation for ethical research practice. When it comes to research publications and reports, being honest and open means above all else avoiding plagiarism—that is, presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment (Hard, Conway, and Moran 2006:1059).

An increasing body of research suggests that plagiarism is a growing problem on college campuses. Jason Stephen, Michael Young, and Thomas Calabrese (2007:243) found in a web survey of self-selected students at two universities that one quarter acknowledged having plagiarized a few sentences (24.7%) or a complete paper (0.3%) in coursework within the past year (many others admitted to other forms of academic dishonesty, such as copying homework). Hard et al. (2006) conducted an anonymous survey in selected classes in one university, with almost all students participating, and found much higher plagiarism rates: 60.6% reported that they had copied "sentences, phrases, paragraphs, tables, figures, or data directly or in slightly modified form from a book, article, or other academic source without using quotation marks or giving proper acknowledgment to the original author or source" (p. 1069) and 39.4% reported that they had "copied information from Internet websites and submitted it as [their] work" (p. 1069).

Plagiarism: Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.

So the plagiarism problem is not just about purchasing term papers—although that is really about as bad as it gets (Broskoske 2005:1); plagiarism is also about what you do with the information you obtain from a literature review or inspection of research reports. And rest assured that this is not only about student papers; it also is about the work of established scholars and social researchers who publish reports that you want to rely on for accurate information. Several noted researchers have been accused of plagiarizing passages that they used in popular books or academic articles; some have admitted to not checking the work of their research assistants, to not keeping track of their sources, or to being unable to retrieve the data they claimed they had analyzed. Serious plagiarism by academic researchers has been identified from New York to Witwatersrand (South Africa), Birmingham (England), and Beijing (Israel 2014:158–159). Whether the cause is cutting corners to meet deadlines or consciously fudging facts, the effect is to undermine the trustworthiness of social research.

What about Wikipedia? With over 5 million Wikipedia articles available in English in 2017 (and 40 million in a total of 293 languages), you can find an article on just about any topic of interest. Of course it is nothing but plagiarism to copy text out of a Wikipedia article and paste it into your paper without proper attribution. But what about using Wikipedia articles as sources for your literature review? Research published in 2005 in the prestigious journal *Nature* found that Wikipedia articles were almost as accurate as those in *Encyclopedia Britannica*—which uses experts to author its articles (Giles 2005). The sheer number of people involved in Wikipedia—contributing text and editing it—generates a self-correcting process that works pretty well even without expert control (Abernathy 2017:42). Yet it's still not appropriate to use Wikipedia or any other encyclopedia as a source in an academic paper or article. The problem is that their articles are distillations of findings from the original sources, and you have no way of knowing how accurately they represent the original findings as they apply to your own research question. That's the hard work you have to do yourself!

Now that you are completing this course in research methods, it's time to think about how to do your part to reduce the prevalence of plagiarism and ensure that you report accurately the findings of any research you conduct. Of course, the first step is to maintain careful procedures for documenting the sources that you rely on for your own research and papers, but you should also think about how best to reduce temptations among others. After all, what people believe about what others do is a strong influence on their own behavior (Hard et al. 2006:1058).

Reviewing the definition of plagiarism and how your discipline's professional association enforces its policies against the practice is an important first step. These definitions and procedures reflect a collective effort to help social scientists maintain standards throughout the discipline. Awareness is the first step (American Sociological Association 1999:19).

Sociologists have an obligation to be familiar with their Code of Ethics, other applicable ethics codes, and their application to sociologists' work. Lack of awareness or misunderstanding of an ethical standard is not, in itself, a defense to a charge of unethical conduct.

The American Sociological Association (ASA)'s (1999) *Code of Ethics* includes an explicit prohibition of plagiarism:

> 14. Plagiarism
>
> (a) In publications, presentations, teaching, practice, and service, sociologists explicitly identify, credit, and reference the author when they take data or material verbatim from another person's written work, whether it is published, unpublished, or electronically available.
>
> (b) In their publications, presentations, teaching, practice, and service, sociologists provide acknowledgment of and reference to the use of others' work, even if the work is not quoted verbatim or paraphrased, and they do not present others' work as their own whether it is published, unpublished, or electronically available. (p. 16)

The next step toward combating the problem and temptation of plagiarism is to keep focused on the goal of social research methods: investigating the social world. If researchers are motivated by a desire to learn about social relations, to understand how people understand society, and to discover why conflicts arise and how they can be prevented, they will be as concerned with the integrity of their research methods as are those, like yourself, who read and use the results of their research. Throughout *Investigating the Social World,* you have been learning how to use research processes and practices that yield valid findings and trustworthy conclusions. Failing to report honestly and openly on the methods used or sources consulted derails progress toward that goal.

It works the same as with cheating in school. When students are motivated only by the desire to "ace" their tests and receive better grades than others, they are more likely to plagiarize and use other illicit means to achieve that goal. Students who seek first to improve their understanding of the subject matter and to engage in the process of learning are less likely to plagiarize sources or cheat on exams (Kohn 2008:6–7). They are also building the foundation for becoming successful researchers who help others understand our social world.

Conclusions

A well-written research article or report requires (to be just a bit melodramatic) blood, sweat, and tears and more time than you will, at first, anticipate. But the process of writing one will help you write the next. And the issues you consider, if you approach your writing critically, will be sure to improve your subsequent research projects and sharpen your evaluations of other investigators' research projects.

Good critical skills are essential when evaluating research reports, whether your own or those produced by others. There are *always* weak points in any research, even published research. It is an indication of strength, not weakness, to recognize areas where one's own research needs to be, or could have been, improved. And it is really not just a question of sharpening your knives and going for the jugular. You need to be able to weigh the strengths and weaknesses of particular research results and to evaluate a study for its contribution to understanding the social world—not whether it gives a definitive answer for all time.

But this is not to say that anything goes. Much research lacks one or more of the three legs of validity—measurement validity, causal validity, or generalizability—and contributes more confusion than understanding about the social world. Top journals generally maintain very high standards, partly because they have good critics in the review process and distinguished editors who make the final acceptance decisions. But some daily newspapers do a poor job of screening, and research-reporting standards in many popular magazines, TV shows, and books are often abysmally poor. Keep your standards high and your views critical when reading research reports, but not so high or so critical that you turn away from studies that make tangible contributions to understanding the social world —even if they don't provide definitive answers. And don't be so intimidated by the need to maintain high standards that you shrink from taking advantage of opportunities to conduct research yourself.

The growth of social science methods from their infancy to adolescence, perhaps to young adulthood, ranks as a key intellectual accomplishment of the 20th century. Opinions about the causes and consequences of homelessness no longer need to depend on the scattered impressions of individuals; criminal justice policies can be shaped by systematic evidence of their effectiveness; and changes in the distribution of poverty and wealth in populations can be identified and charted. Employee productivity, neighborhood cohesion, and societal conflict may each be linked to individual psychological processes and to international economic strains.

Of course, social research methods are no more useful than the commitment of researchers to their proper application. Research methods, like all knowledge, can be used poorly or well, for good purposes or bad, when appropriate or not. A claim that a belief is based on social science research in itself provides no extra credibility. As you have learned throughout this book, we must first learn which methods were used, how they were applied, and whether interpretations square with the evidence. To investigate the social world, we must keep in mind the lessons of research methods. Doing so will help us build a better social world in the 21st century.

SAGE edge

Want a better grade? Get the tools you need to sharpen your study skills. Access practice quizzes, eFlashcards, video, and multimedia at edge.sagepub.com/schutt9e

Key Terms

Back matter 605
Data visualization 596
Front matter 605
Plagiarism 611
Publication bias 610
Reverse outlining 593

Highlights

- Research reports should be evaluated systematically, using the review guide in Appendix A and considering the interrelations between the design elements.
- Data visualization should seek to be trustworthy, accessible, and elegant.
- All write-ups of research should be revised several times and critiqued by others before being presented in final form. Reverse outlining can help in this process.
- Different types of reports typically pose different problems. Applied research reports are constrained by the expectations of the research sponsor; an advisory committee from the applied setting can help avoid problems. Journal articles must pass a peer review by other social scientists and are often much improved in the process.
- Research reports should include an introductory statement of the research problem, a literature review, a methodology section, a findings section with pertinent data displays, and a conclusions section that identifies any weaknesses in the research design and points out implications for future research and theorizing. This basic report format should be modified according to the needs of a particular audience.
- The central ethical concern in research reporting is to be honest. This honesty should include providing a truthful accounting of how the research was carried out, maintaining a full record about the project, using appropriate statistics and graphs, acknowledging the research sponsors, and being sensitive to the perspectives of coauthors.
- Credit must be given where credit is due. The contributions of persons and organizations to a research project must be acknowledged in research reports.
- Publication bias results from the greater likelihood of acceptance of articles reporting confirmation of hypothesized effects.
- Plagiarism is a grievous violation of scholarly ethics. All direct quotes or paraphrased material from another author’s work must be appropriately cited.
- Social scientists are obligated to evaluate the credibility of information obtained from any source before using it in their research reports.

Discussion Questions

1. A good place to start developing your critical skills would be with one of the articles reviewed in this chapter. Try reading one, and fill in the answers to the article review questions that I did not cover (see Appendix A). Do you agree with my answers to the other questions? Could you add some points to my critique or to the lessons on research design that I drew from these critiques?
2. Read the journal article "Marital Disruption and Depression in a Community Sample," by Aseltine and Kessler, in the September 1993 issue of *Journal of Health and Social Behavior.* How effective is the article in conveying the design and findings of the research? Could the article's organization be improved at all? Are there bases for disagreement about the interpretation of the findings? Did reading the full article increase your opinion of its value?
3. Rate four journal articles on the study site, at edge.sagepub.com/schutt9e, for the overall quality of the research and for the effectiveness of the writing and data displays. Discuss how each could have been improved.

Practice Exercises

1. Call a local social or health service administrator or a criminal justice official, and arrange for an interview. Ask the official about his or her experience with applied research reports and conclusions about the value of social research and the best techniques for reporting to practitioners.
2. Interview a professor who has written a research article for publication based on original data. Ask the professor to describe his or her experiences while writing the article. Review the decisions made in reporting the research design, developing a plan for data analysis, and writing up findings and conclusions. What issues seemed to be most challenging in this process?
3. Complete the interactive exercises on reporting research on the book's study site, at edge.sagepub.com/schutt9e.

Ethics Questions

1. Plagiarism is no joke. What are the regulations on plagiarism in class papers at your school? What do you think the ideal policy would be? Should this policy account for cultural differences in teaching practices and learning styles? Do you think this ideal policy is likely to be implemented? Why or why not? Based on your experiences, do you believe that most student plagiarism is the result of misunderstanding about proper citation practices, or is it the result of dishonesty? Do you think that students who plagiarize while in school are less likely to be honest as social researchers?
2. Full disclosure of funding sources and paid consulting and other business relationships is now required by most journals. Should researchers publishing in social science journals also be required to fully disclose all sources of funding, including receipt of payment for research done as a consultant? Should full disclosure of all previous funding sources be required in each published article? Write a short justification of the regulations you propose.
3. What should be done about the problem of publication bias? Should there be a *Journal of Null Findings?* Does the requirement of registration at clinicaltrials.gov solve the problem for medical research? Should that requirement be extended to all social science experiments? What are the pros and cons of “full disclosure” about all hypothesized effects that have been tested in research projects using experimental designs?

Web Exercises

1. Go to the National Science Foundation's (NSF) Sociology Program website at www.nsf.gov/funding/pgm_summ.jsp?pims_id=5369. What are the components that the NSF's Sociology Program looks for in a proposed piece of research? Examine the requirements for an NSF proposal at https://www.nsf.gov/pubs/policydocs/pappg17_1/nsf17_1.pdf. Now review the "Project Description" section (p. 18) and write a brief description along these lines for a research investigation of your choice.
2. The National Academy of Sciences wrote a lengthy report on ethics issues in scientific research. Visit the site and read the free executive summary. Go to www.nap.edu/catalog.php?record_id=10430 and click on "Download Free PDF" (you can download as a "guest"). Summarize the information and guidelines in the report.
3. Using the web, find five different examples of social science research projects that have been completed. Briefly describe each. How does each differ in its approach to reporting the research results? To whom do you think the author(s) of each are "reporting" (i.e., who is the audience)? How do you think the predicted audience has helped shape the author's approach to reporting the results? Be sure to note the websites at which you located each of your five examples.

Video Interview Questions

Listen to my interview for Chapter 16 at edge.sagepub.com/schutt9e.

1. What were our primary research findings?
2. What changes did the Women's Health Network implement in light of the research findings?

SPSS Exercises

1. Review the output you have generated in previous SPSS exercises. Select the distributions, statistics, and crosstabs that you believe provide a coherent and interesting picture of support for capital punishment in the United States. Prepare these data displays using the graphic techniques presented in this chapter and number them (Figure 1, Figure 2, etc.).
2. Write a short report based on the analyses you conducted for the SPSS exercises throughout this book, including the data displays you have just prepared. Include in your report a brief introduction and literature review (you might use the articles I referred to in the SPSS exercises for Chapter 2). In a short methods section, review the basic methods used in the GSS 2016, and list the variables you have used for the analysis.
3. In your conclusions section, include some suggestions for additional research on support for capital punishment.

Developing a Research Proposal

Now it's time to bring all the elements of your proposal together (Exhibit 3.10, #19 to #23).

1. Organize the proposal material you wrote for previous chapters in a logical order. Select what you feel is the strongest research method (see Chapters 7, 8, and 10) as your primary method.
2. Consider a complex research design and include at least one additional method of data collection (see Chapters 12–15).
3. Specify a data analysis plan (Chapters 9 and/or 11).
4. Rewrite the entire proposal, adding an introduction. Also add sections that outline a budget, and state the limitations of your study.
5. Review the proposal with the "Decisions in Research" checklist (see Exhibit 3.10). Answer each question (or edit your previous answers), and justify your decision at each checkpoint.

Appendix A Questions to Ask About a Research Article

1. What is the basic research question, or problem? Try to state it in just one sentence. (Chapter 2)
2. Is the purpose of the study descriptive, exploratory, explanatory, or evaluative? Did the study have more than one purpose? (Chapter 1)
3. Was a theoretical framework presented? What was it? Did it seem appropriate for the research question addressed? Can you think of a different theoretical perspective that might have been used? What philosophy guides the research? Is this philosophy appropriate to the research question? (Chapters 1, 2)
4. What prior literature was reviewed? Was it relevant to the research problem? Was it relevant to the theoretical framework? Does the literature review appear to be adequate? Are you aware of (or can you locate) any important studies that have been omitted? (Chapter 2)
5. What features identified the study as deductive or inductive? Do you need additional information in any areas to evaluate the study? (Chapters 1, 2)
6. Did the study seem consistent with current ethical standards? Were any trade-offs made between different ethical guidelines? Was an appropriate balance struck between adherence to ethical standards and use of the most rigorous scientific practices? (Chapter 3)
7. Were any hypotheses stated? Were these hypotheses justified adequately in terms of the theoretical framework and in terms of prior research? (Chapter 2)
8. What were the independent and dependent variables in the hypothesis or hypotheses? Did these variables reflect the theoretical concepts as intended? What direction of association was hypothesized? Were any other variables identified as potentially important? (Chapter 2)
9. What were the major concepts in the research? How, and how clearly, were they defined? Were some concepts treated as unidimensional that you think might best be thought of as multidimensional? (Chapter 4)
10. Did the instruments used, the measures of the variables, seem valid and reliable? How did the authors attempt to establish this? Could any more have been done in the study to establish measurement validity? (Chapter 4)
11. Was a sample or the entire population of elements used in the study? What type of sample was selected? Was a probability sampling method used? Did the authors think the sample was generally representative of the population from which it was drawn? Do you? How would you evaluate the likely generalizability of the findings to other populations? (Chapter 5)
12. Was the response rate or participation rate reported? Does it appear likely that those who did not respond or participate were markedly different from those who did

participate? Why or why not? Did the author(s) adequately discuss this issue? (Chapter 5)

13. What were the units of analysis? Were they appropriate for the research question? If groups were the units of analysis, were any statements made at any point that are open to the ecological fallacy? If individuals were the units of analysis, were any statements made at any point that suggest reductionist reasoning? (Chapter 6)
14. Was the study design cross-sectional or longitudinal, or did it use both types of data? If the design was longitudinal, what type of longitudinal design was it? Could the longitudinal design have been improved in any way, as by collecting panel data rather than trend data, or by decreasing the dropout rate in a panel design? If cross-sectional data were used, could the research question have been addressed more effectively with longitudinal data? (Chapter 6)
15. Were any causal assertions made or implied in the hypotheses or in subsequent discussions? What approach was used to demonstrate the existence of causal effects? Were all three criteria and two cautions for establishing causal relationships addressed? What, if any, variables were controlled in the analysis to reduce the risk of spurious relationships? Should any other variables have been measured and controlled? How satisfied are you with the internal validity of the conclusions? What about external validity? (Chapter 6)
16. Which type of research design was used: experimental, survey, participant observation, historical comparative, or some other? How well was this design suited to the research question posed and the specific hypotheses tested, if any? Why do you suppose the author(s) chose this particular design? How was the design modified in response to research constraints? How was it modified to take advantage of research opportunities? (Chapters 7, 8, 10, 13, 14, 15)
17. Were mixed methods used? What methods were combined and how can they be distinguished in priority and sequence? In what ways did the combination of qualitative and quantitative data enrich the study's value? (Chapter 12)
18. Was this an evaluation research project? If so, which type of evaluation was it? Which design alternatives did it use? (Chapter 13)
19. Was a secondary data analysis design used? If so, what were the advantages and disadvantages of using data collected in another project? Were Big Data analyzed? If so, did the methods raise any ethical alarms? (Chapter 14)
20. Was a historical comparative design or a content analysis used? Which type was it? Did the authors address problems resulting from the use of historical or cross-national data? (Chapter 15)
21. Was any attention given to social context and subjective meanings? If so, what did this add? If not, would it have improved the study? Explain. (Chapter 10)
22. Summarize the findings. How clearly were statistical and/or qualitative data presented and discussed? Were the results substantively important? If meta-analysis was used, was the selection of studies comprehensive and were clear generalizations possible from the population of studies? (Chapters 9, 11)

23. Did the author(s) adequately represent the findings in the discussion or conclusions sections? Were conclusions well grounded in the findings? Are any other interpretations possible? (Chapter 16)
24. Compare the study to others addressing the same research question. Did the study yield additional insights? In what ways was the study design more or less adequate than the design of previous research? (Chapters 2, 16)
25. What additional research questions and hypotheses are suggested by the study's results? What light did the study shed on the theoretical framework used? On social policy questions? (Chapters 2, 13, 16)

Appendix B How to Read a Research Article

The discussions of research articles throughout the text may provide all the guidance you need to read and critique research on your own. But reading about an article in bits and pieces to learn about particular methodologies is not quite the same as reading an article in its entirety to learn what the researcher found out. The goal of this appendix is to walk you through an entire research article, answering the review questions introduced in Appendix A. Of course, this is only one article and our "walk" will take different turns than would a review of other articles, but after this review you should feel more confident when reading other research articles on your own.

We will use for this example an article by Seth Abrutyn and Anna S. Mueller (2014) on suicidal behavior among adolescents, reprinted on pages 624 to 640 of this appendix. It focuses on a topic of great social concern and of key importance in social theory. Moreover, it is a solid piece of research published by a top SAGE journal, the American Sociological Association's *American Sociological Review.*

I have reproduced below each of the article review questions from Appendix A, followed by my answers to them. After each question, I indicate the chapter where the question was discussed and after each answer, I cite the article page or pages that I am referring to. You can also follow my review by reading through the article itself and noting my comments.

1. *What is the basic research question, or problem? Try to state it in just one sentence.* (Chapter 2)

   Abrutyn and Mueller present an overall research problem and then four specific research questions. They define their research problem by stating, "We investigate the role suicide suggestion plays in the suicide process, independent of other measures of social integration and psychological well-being" (p. 212). They summarize their research questions as being "the critical questions of how, when, and for whom does suggestion matter?" The four specific research questions highlight

   Source: *American Sociological Review*, 2014, Vol. 79(2):211–227.

   > Four major gaps in the literature: (1) whether suicide suggestion is associated with the development of suicidal thoughts among individuals who reported no suicidal thoughts at the time a role model attempted suicide; (2) whether the effects of suicide suggestion fade with time; (3) whether the relationship between the role model and respondent matters; and (4) whether there are differences between boys and girls. (p. 212)

Before this point, the authors focus on this research question by highlighting the apparent paradox that the social integration that Émile Durkheim assumed helped protect individuals from suicide can instead spread suicidality.

2. *Is the purpose of the study descriptive, exploratory, explanatory, or evaluative? Did the study have more than one purpose?* (Chapter 1)

   The study's primary purpose is explanatory because the authors conclude each section of their literature review with a possible explanation of influences on risk of suicidality. For example, the section on "gender differences" concludes with a summary statement that highlights the goal of explanation:

   > These findings suggest girls may be more susceptible than boys to role models' suicide attempts. (p. 215)

   There is also a descriptive element in the authors' framing of the research because they indicate their strategy includes "examining the development of suicidal behaviors in a sample of youth with no suicidal behaviors at Time I" (p. 215).

   Of course, the authors also present descriptive statistics for their key variables (Table 1, p. 217).

3. *Was a theoretical framework presented? What was it? Did it seem appropriate for the research question addressed? Can you think of a different theoretical perspective that might have been used? What philosophy guides the research? Is this philosophy appropriate to the research question?* (Chapters 1, 2)

   Abrutyn and Mueller's overarching theoretical framework for this research is Durkheim's classic theory of suicide and its emphasis on the protective value of social integration (pp. 212–215). The article begins and ends by discussing Durkheim's theory, and it introduces the concept of suicide suggestion as important for sociologists because of the "apparent contradiction" it involves with Durkheim's theory. The literature review is focused on theorizing and research about suicide suggestion and so is very appropriate for the research questions addressed. Some connections are made to identity theory, which provides a somewhat different theoretical perspective that is more appropriate for some of the influences tested. The researchers follow a positivist research philosophy as they seek to understand social processes in the social world, rather than how people make sense of their experiences. In this study, the focus on suicide "suggestion" certainly raises a question about meaning, but their methods use standard measures to identify variation in this phenomenon, rather than intensive interviews or observations to discover how adolescents construct the experience of a friend's suicide. We can conclude that the positivist philosophy guiding the research is appropriate to the research question, while realizing that a researcher guided by a constructivist philosophy could have studied the same phenomenon but with different research methods and a somewhat different research question.

4. *What prior literature was reviewed? Was it relevant to the research problem? Was it relevant to the theoretical framework? Does the literature review appear to be adequate? Are you aware of (or can you locate) any important studies that have been omitted?* (Chapter 2)

   Abrutyn and Mueller review literature from the article's first page until the methods section (pp. 212–215). It is all very relevant to the general theoretical framework problem, and there is a section focused on each of the four specific research questions. In the first few paragraphs, several studies are mentioned that draw attention to the importance of suicide suggestion and thus the potential negative effects of social ties on suicidality (pp. 211–212). Subsequent sections in the literature review focus on prior research about suicidality and the effects of media, role models, recency ("temporal limitations"), family versus friends, and gender. The review provides an adequate foundation for expecting these effects. I leave it to you to find out whether any important studies were omitted.
5. *What features identified the study as deductive or inductive? Do you need additional information in any areas to evaluate the study?* (Chapters 1,2)

   The study clearly involves a test of ideas against empirical reality as much as that reality could be measured; it was carried out systematically and with a deductive design. Because the authors used an available data set, others can easily obtain the complete documentation for the study and try to replicate the authors' findings. The authors explicitly note and challenge assumptions made by many other researchers using Durkheim's theory of social integration and suicide (p. 211). They aim clearly to build social theory and encourage others to build on their findings: "this study is not without its limitations" (p. 224). The study thus seems to exemplify adherence to the logic of deductive research and to be very replicable.
6. *Did the study seem consistent with current ethical standards? Were any trade-offs made between different ethical guidelines? Was an appropriate balance struck between adherence to ethical standards and use of the most rigorous scientific practices?* (Chapter 3)

   Abrutyn and Mueller use survey data collected by others and so encounter no ethical problems in their treatment of human subjects. The reporting seems honest and open. Although the research should help inform social policy, the authors' explicit focus is on how their research can inform social theory. This is quite appropriate for research reported in a scientific journal, so there are no particular ethical problems raised about the uses to which the research is put. The original survey used by the authors does not appear at all likely to have violated any ethical guidelines concerning the treatment of human subjects, although it would be necessary to inspect the original research report to evaluate this.

7. *Were any hypotheses stated? Were these hypotheses justified adequately in terms of the theoretical framework and in terms of prior research?* (Chapter 2)

   Although they do not explicitly label their predictions as hypotheses, Abrutyn and Mueller carefully specify their independent and dependent variables and link them to their five specific research questions. Each one is justified and related to prior research in the theoretical background section (pp. 212–215).

8. *What were the independent and dependent variables in the hypothesis or hypotheses? Did these variables reflect the theoretical concepts as intended? What direction of association was hypothesized? Were any other variables identified as potentially important?* (Chapter 2)

   Independent and dependent variables are identified explicitly in the measurement section and several control variables are specified as important (pp. 216–218). Independent variables are friend suicide attempt, family suicide attempt, family integration scale, friends care, and religious attendance. Although it is not stated explicitly as an independent variable, the authors identify another independent variable, recency of others' suicide, by distinguishing the survey follow-up wave at which that event occurred and the wave at which the dependent variable is measured. Additional variables controlled as known risk factors for suicide are same-sex attraction and emotional distress. Several demographic and personal characteristics are also used as controls: grade point average, family structure, race/ethnicity, parents' education, and age.

9. *What were the major concepts in the research? How, and how clearly, were they defined? Were some concepts treated as unidimensional that you think might best be thought of as multidimensional?* (Chapter 4)

   The key concept in the research is that of "suicide suggestion." It is defined in the article's second sentence and distinguished from the parallel concept of "social integration" that Durkheim emphasized (p. 211). Two dimensions of this key concept are then distinguished (emotional—suicidality and behavioral—suicides) and the concept of suicidality is discussed at length in a section on the spread of suicide (p. 212). Other important concepts in the research are personal role models, similarity between individuals and role models, type of role model (family versus friends), temporal limits, and gender. They are each elaborated in the separate sections of the literature review. Several related concepts are mentioned in the course of discussing others, including significant others, reality of role model (p. 212), media exposure, depression, social similarity of friends (p. 214), suggestibility and network diffusion (p. 215), family integration and care by others (p. 216), and religious attendance (p. 217).

10. *Did the instruments used, the measures of the variables, seem valid and reliable? How did the authors attempt to establish this? Could any more have been done in the study to*

establish measurement validity? (Chapter 4)

The measures of the dependent variables, suicidal ideation and suicide attempts, were based on answers to single questions (p. 216). The wording of these questions is quite straightforward, but no information is provided about their reliability or validity. This can be seen as a weakness of the article, although Abrutyn and Miller do note in the limitations section that they do not know how stated thoughts or intentions were related to actual suicidal behavior (p. 225). The same single-question approach was used to measure the independent variables, friend suicide attempt and family suicide attempt (p. 216), again without any information on reliability or validity. The authors report that family integration was measured with a four-item scale, for which they report interitem reliability, and relations with friends, religious attendance, and same-sex attraction were assessed with just single questions (pp. 216–217). Abrutyn and Mueller measured emotional distress with an abridged version of the widely used Center for Epidemiological Studies Depression (CES-D) scale (pp. 217–218). They mention the interitem reliability of the CES-D in their data ("Cronbach's alpha = .873") but do not discuss its validity; because the scale has been used in so many other studies and has been evaluated so many times for its reliability and validity, and because it does not play a central role in this new research, it is reasonable that the authors do not discuss it further. Overall, the study can be described as relatively weak in information provided on measurement reliability and validity.

11. *Was a sample or the entire population of elements used in the study? What type of sample was selected? Was a probability sampling method used? Did the authors think the sample was generally representative of the population from which it was drawn? Do you? How would you evaluate the likely generalizability of the findings to other populations?* (Chapter 5)

The sample was a national random (probability) sample of adolescents at three time points. Called the "Add Health" study, the original researchers used a two-stage cluster sampling design, in which first all schools in the United States containing an 11th grade were sampled, with stratification by region, urbanicity, school type, ethnic composition, and size. A nationally representative subsample of students was then interviewed from these schools (n = 20,745). Students (and graduates) from the subsample were then reinterviewed in two more waves, with Wave II 1 to 2 years after Wave I and Wave III another 5 to 6 years after Wave II. A total of 10,828 respondents were interviewed in all three waves (p. 215). Abrutyn and Mueller attempt to determine whether suicidality is influenced among adolescents who have not previously thought of suicide by processes of contagion, so they limit their sample further to those who reported no suicidal thoughts or attempts in Wave I (p. 216). The authors identify their sample explicitly as representative of the national population of adolescents in school in this age range. Do you think the findings could be generalized to adolescents who had dropped out of school, or to other countries with different cultural values about suicide?

12. *Was the response rate or participation rate reported? Does it appear likely that those who did not respond or participate were markedly different from those who did participate? Why or why not? Did the author(s) adequately discuss this issue?* (Chapter 5)

    The number of cases is identified at each wave, and the consequences of exclusion criteria applied are specified, but the response rate is not stated. Readers are referred for details to the original research report on the Add Health survey (p. 215). The authors do evaluate the possibility that the exclusion of adolescents who reported suicidal ideation in Wave I could have biased their sample, but they suggest this is unlikely because average levels of emotional distress and demographic variables are similar whether these cases are excluded or not (p. 216). There does not seem to be a serious problem here, but it would have been helpful to have had details about the response rate in the article, instead of just in a separate report (albeit one that is available online). More consideration of possibilities for sample bias could also have led to greater confidence in generalizations from this sample.
13. *What were the units of analysis? Were they appropriate for the research question? If groups were the units of analysis, were any statements made at any point that are open to the ecological fallacy? If individuals were the units of analysis, were any statements made at any point that suggest reductionist reasoning?* (Chapter 6)

    The survey sampled adolescents within schools, so individuals were the units of analysis (pp. 215–216). The focus was on the behavior of individuals, so this is certainly appropriate. However, it is possible that the process of suicide contagion could differ between different schools, so the authors could have added a great deal to their study by also using schools as the units of analysis and determining whether there were some distinctive characteristics of schools in which contagion was more likely. Therefore, the individual-level analysis could obscure some group-level processes and thus lead to some reductionist reasoning.
14. *Was the study design cross-sectional or longitudinal, or did it use both types of data? If the design was longitudinal, what type of longitudinal design was it? Could the longitudinal design have been improved in any way, as by collecting panel data rather than trend data, or by decreasing the dropout rate in a panel design? If cross-sectional data were used, could the research question have been addressed more effectively with longitudinal data?* (Chapter 6)

    The study used a longitudinal panel design, although the sample at Wave II was limited to those adolescents who had not already graduated (p. 215). The reduction in the sample size by about half from the first wave to the third follow-up 5 to 6 years later is typical for a panel design; it is not possible to consider whether procedures could have been improved without knowing more about the details contained in the original research report.
15. *Were any causal assertions made or implied in the hypotheses or in subsequent discussions?*

What approach was used to demonstrate the existence of causal effects? Were all three criteria and two cautions for establishing causal relationships addressed? What, if any, variables were controlled in the analysis to reduce the risk of spurious relationships? Should any other variables have been measured and controlled? How satisfied are you with the internal validity of the conclusions? What about external validity? (Chapter 6)

Causal assertions are implied in the predictions about "how exposure to suicidal behaviors shapes adolescent suicidality" (p. 215). A nomothetic approach to causation is used and each of the criteria for establishing causal effects is addressed: association (by checking for an association between the independent and dependent variables), time order (by using a longitudinal design that establishes clearly that the precipitating cause occurred before the "contagion"), and nonspuriousness (by controlling for variables that could have created a spurious association). The variables controlled included family integration, closeness to friends, religious attendance, same-sex attraction, and emotional distress, as well as several personal and demographic characteristics. The combination of the longitudinal design and the breadth of the variables controlled increases confidence in the internal validity of the conclusions, but because there was not random assignment to experiencing a friend or family member's suicide (an ethical and practical impossibility), we cannot be completely confident that adolescents exposed to suicide did not differ from their peers in some unmeasured risk factor. There is little basis for evaluating external validity.

16. *Which type of research design was used: experimental, survey, participant observation, historical comparative, or some other? How well was this design suited to the research question posed and the specific hypotheses tested, if any? Why do you suppose the author(s) chose this particular design? How was the design modified in response to research constraints? How was it modified to take advantage of research opportunities?* (Chapters 7, 8, 10, 13, 14, 15)

    Survey research was the method used in the Add Health study, which generated the data set used in this secondary data analysis project. Survey research seems appropriate for the research questions posed, although only because Add Health used a longitudinal panel design. The survey design was somewhat modified after the fact for this analysis by eliminating respondents who had already experienced a suicide of a friend or family member at Wave I (p. 215).

17. *Were mixed methods used? What methods were combined and how can they be distinguished in terms of priority and sequence? In what ways did the combination of qualitative and quantitative data enrich the study's value?* (Chapter 12)

    Mixed methods were not used. The analysis is entirely of quantitative data. The original study could have been enriched for the purposes of addressing suicide contagion by adding qualitative interviews of students who had been exposed to a suicide, as the authors note (p. 223).

18. *Was this an evaluation research project? If so, which type of evaluation was it? Which design alternatives did it use?* (Chapter 13)

    This study did not use an evaluation research design. The issues on which it focused might profitably be studied in some evaluations of adolescent suicide prevention programs.

19. *Was a secondary data analysis design used? If so, what were the advantages and disadvantages of using data collected in another project? Were Big Data analyzed? If so, did the methods raise any ethical alarms?* (Chapter 14)

    This article reported a secondary analysis of the Add Health survey data (p. 215). This analysis of the previously collected data allowed the authors to conduct a very careful and fruitful analysis of a research question that had not previously been answered with these data, without having to secure funds for conducting such an ambitious longitudinal survey themselves. However, the result was that they could not include more extensive questions about suicidality and friends' and family members' suicides as they probably would have done if they had designed the primary data collection themselves. This standard survey data set would not qualify as Big Data.

20. *Was a historical comparative design or a content analysis used? Which type was it? Did the authors address problems resulting from the use of historical or cross-national data?* (Chapter 15)

    This study did not use any type of historical or comparative design. It is interesting to consider how the findings might have differed if comparisons with other cultures or to earlier times had been made. The authors include in their literature review content analyses of the reporting of suicide as part of investigating media effects on suicide, but they did not take this approach themselves (p. 213).

21. *Was any attention given to social context and subjective meanings? If so, what did this add? If not, would it have improved the study? Explain.* (Chapter 10)

    Social context and subjective meanings are really at the heart of the phenomenon of suicide contagion, but the Add Health researchers did not add a qualitative component to their research or adopt a constructivist philosophy to focus more attention on these issues. A participant observation study of suicide contagion in high schools could yield great insights, and a researcher with a constructivist philosophy who focused on how adolescents make sense of others' suicides would add another dimension to understanding this phenomenon.

22. *Summarize the findings. How clearly were statistical or qualitative data presented and discussed? Were the results substantively important?* (Chapters 9, 11)

    Statistical data are presented clearly in two tables. The first table describes the sample at each wave in terms of the key variables for the study. The second table presents the results of a multivariate analysis of the data using a technique called logistic regression. The authors use this technique to test for the associations between their dependent and independent variables over time, while controlling for other variables. The results seem substantively important. In the authors' own words,

> For adolescents, ties do bind, but whether these ties integrate adolescents into society, with positive repercussions for their emotional well-being, or whether they promote feelings of alienation, depends in part on the qualities embedded in those ties. (p. 225)

23. *Did the author(s) adequately represent the findings in the discussion or conclusions sections? Were conclusions well grounded in the findings? Are any other interpretations possible?* (Chapter 16)

    The findings are well represented in the discussion and conclusions section, with a limitations section that strikes an appropriate note of caution (pp. 222–225). Interesting conjectures are presented in the discussion of gender differences and differences in the apparent effects of different types of role model. The conclusions section makes explicit connections to the initial questions posed about Durkheim's theorizing. You might want to consider what other interpretations of the findings might be possible. Remember that other interpretations always are possible for particular findings—it is a question of the weight of the evidence, the persuasiveness of the theory used, and the consistency of the findings with other research.
24. *Compare the study to others addressing the same research question. Did the study yield additional insights? In what ways was the study design more or less adequate than the design of previous research?* (Chapters 2, 16)

    Summaries of prior research in the literature review suggest that Abrutyn and Mueller have added new insights to the literature on suicide contagion by overcoming limitations in previous research designs. The use of a longitudinal panel design was more adequate than much previous research using cross-sectional designs.
25. *What additional research questions and hypotheses are suggested by the study's results? What light did the study shed on the theoretical framework used? On social policy questions?* (Chapters 2, 13, 16)

    Perhaps the most obvious research question suggested by the study's results is that of whether social integration of some type can have protective effects as predicted by Durkheim even at the same time that other social connections heighten the risk of suicide due to a process of social contagion. Research designed to answer this question could lead to an overarching theoretical framework encompassing both the protective benefits of social ties that Durkheim identified and the risk that social ties create a process of social contagion. If the focus is on understanding the process of social contagion, it is clear that Abrutyn and Mueller's research has made an important contribution. The authors highlight some policy implications in their conclusions (p. 225).

Are Suicidal Behaviors Contagious in Adolescence? Using Longitudinal Data to Examine Suicide Suggestion

Seth Abrutyn[a] and Anna S. Mueller[a]

American Sociological Review 2014, Vol. 79(2) 211–227

© American Sociological Association 2014

DOI: 10.1177/0003122413519445

http://asr.sagepub.com

Abstract

Durkheim argued that strong social relationships protect individuals from suicide. We posit, however, that strong social relationships also have the potential to increase individuals' vulnerability when they expose people to suicidality. Using three waves of data from the National Longitudinal Study of Adolescent Health, we evaluate whether new suicidal thoughts and attempts are in part responses to exposure to role models' suicide attempts, specifically friends and family. We find that role models' suicide attempts do in fact trigger new suicidal thoughts, and in some cases attempts, even after significant controls are introduced. Moreover, we find these effects fade with time, girls are more vulnerable to them than boys, and the relationship to the role model—for teenagers at least—matters. Friends appear to be more salient role models for both boys and girls. Our findings suggest that exposure to suicidal behaviors in significant others may teach individuals new ways to deal with emotional distress, namely by becoming suicidal. This reinforces the idea that the structure—and content—of social networks conditions their role in preventing suicidality. Social ties can be conduits of not just social support, but also antisocial behaviors, like suicidality.

Keywords

suicide, social networks, suicide suggestion, Durkheim, gender, Add Health

Understanding suicide has been essential to the sociological enterprise since Durkheim ([1897] 1951) wrote his famous monograph, arguing that groups that integrated and (morally) regulated their members offered protective benefits against suicide. Durkheimian mechanisms remain highly relevant (cf. Maimon and Kuhl 2008; Pescosolido and Georgianna 1989; Thorlindsson and Bjarnason 1998), but emphasis on *suicide suggestion,* or the effect a role model's suicidal behavior has on an observer's suicidality, has become increasingly essential to the sociological understanding of suicide (e.g., Gould 2001; Phillips 1974; Stack 2003, 2009).

Whereas Durkheim assumed that social integration protected individuals, suicide suggestion demonstrates that suicidality can spread through the very ties that Durkheim theorized as protective. This apparent contradiction is not such a problem for modern interpretations of Durkheim's theory that focus on the structure of social ties themselves, and how the networks individuals are embedded within produce the protective benefits Durkheim observed (Bearman 1991; Pescosolido 1990; Wray, Colen, and Pescosolido 2011). It is possible to imagine social ties as capable of both social support and social harm (Baller and Richardson 2009; Haynie 2001; Pescosolido 1990). Durkheim was right that collective solidarity is often protective, but we argue that the behaviors, values, and emotions embedded in network ties must be elaborated to truly understand how social relationships shape individuals' life chances.[1] This subtle shift provides an opportunity to integrate two equally important, but often unnecesarrily separate, realms in the sociology of suicide: the literature on suicide suggestion and the literature on social integration.

The existing literature on suicide suggestion demonstrates that concern over the emotions (suicidality) and behaviors (suicides) embedded in social networks is warranted. Suicides often occur in clusters, with spikes in suicide rates following media coverage of suicides (Stack 2003, 2005, 2009), so much so that a group of public health agencies (including the National Institute of Mental Health [NIMH]) issued guidelines for how the media should report on suicides so as to limit their spread (Suicide Prevention Resource Center [SPRC] 2013). Less research has examined how suicides spread through personal role models, but studies show a robust association between a friend's (and sometimes family member's) suicidal behavior and that of the person exposed to it (Bearman and Moody 2004; Bjarnason 1994; Liu 2006; Niederkrotenthaler et al. 2012; Thorlindsson and Bjarnason 1998). However, these studies often fail to address the critical questions of how, when, and for whom does suggestion matter?

With this study, we employ three waves of data from the National Longitudinal Study of

Adolescent Health to examine these questions. By using longitudinal data rich in measures of adolescent life, we investigate the role suicide suggestion plays in the suicide process, independent of other measures of social integration and psychological well-being. We tease out nuances related to the harmful side of social integration by shedding light on four major gaps in the literature: (1) whether suicide suggestion is associated with the development of suicidal thoughts among individuals who reported no suicidal thoughts at the time a role model attempted suicide; (2) whether the effects of suicide suggestion fade with time; (3) whether the relationship between the role model and respondent matters; and (4) whether there are differences between boys and girls.

Theoretical Background

The Spread of Suicide

Beginning with Phillips's (1974) groundbreaking work, suicide suggestion studies typically examine (1) the association between celebrity suicides and national and local suicide rates (Gould 2001; Stack 2003, 2005), (2) the association between fictionalized media suicides and national and local rates (e.g., Stack 2009), and (3) the apparent geographic and temporal clustering of suicides (e.g., Baller and Richardson 2002; Gould, Wallenstein, and Kleinman 1990). A few studies have also investigated the effect a role model's suicidal behavior has on friends or family members exposed to it. The logic of these studies is predicated on social psychological assumptions. Significant others or persons labeled as members of a reference group with whom we identify are far more likely to influence and shape behavior than are nonsignificant others or outsiders (Turner 2010). Additionally, direct ties infused with socioemotional meanings can act as conduits for the spread of behavioral norms (Goffman 1959) and positive *and* negative affect, which motivate the reproduction of these behavioral norms (Lawler 2006).

Suicide suggestion and the media. In a comprehensive review of the suicide suggestion literature, Stack (2005:121) estimates that about one-third of suicide cases in the United States involve "suicidal behavior following the dissemintation of a suicidal model in the media." Models may be real celebrities like Marilyn Monroe or fictionalized characters such as those found in popular novels or television shows. The length of exposure and the status of the role model appear to matter: on average, publicized celebrity suicides produce a 2.51 percent spike in aggregate rates, whereas Marilyn Monroe's suicide, a high status and highly publicized suicide, was followed by a 13 percent spike in the U.S. suicide rate (Phillips 1974; Stack 2003). The evidence concerning effects of fictionalized suicides, such as those found occasionally in television series (Schmidtke and Hafner 1988), is less consistent (e.g., Niederkrontenthaler and Sonneck 2007), but a recent metaanalysis found youths are particularly at risk of suicide suggestion via fictional suicides (Stack 2009).

Spikes following celebrity suicides are confined geographically to the subpopulation exposed to the suicide—for example, local newspapers should only affect their readership, whereas nationally televised shows should reach more people. Furthermore, research shows that temporal effects of media exposure vary to some degree, typically ranging from two weeks to a month (Phillips 1974; Stack 1987). To date, these studies have had a difficult time determining whether suggestion plays a role above and beyond individuals' personal circumstances: finding an association between media and suicide rates "does not necessarily identify [suggestion] as the underlying mechanism" (Gould et al. 1990:76). If suicide suggestion plays a role in the suicide process, the question is: does it have an effect above and beyond other risk factors for suicide, such as suicidal thoughts or depression prior to exposure to media coverage of a suicide?

Suicide suggestion via personal role models. Like media exposure suggestion studies, studies of personal role models focus on demonstrating a link between a role model's and the exposed individual's suicidal behaviors. The majority of these studies focus on adolescent suicide, perhaps because adolescent suicide has tripled since the 1950s and thus represents a serious public health problem (NIMH 2003). Adolescents may also be particularly vulnerable to suicide suggestion: adolescents are particularly socially conscious—social status and social relationships are a major focus of their daily lives. Moreover, teenagers are greatly influenced by their peers' values and behaviors (Giordano 2003), which may increase their vulnerability to suicide epidemics. Finally, adolescents are unique in that their sense of self is still forming, so they are more malleable than adults (Crosnoe 2000; Crosnoe and Johnson 2011). Any insights into factors contributing to the development of suicidality are thus crucial to teen suicide prevention.

Generally, studies of personal role models show that having a friend or family member exhibit suicidal behavior is positively associated with an exposed adolescent's own suicidality (Bjarnason and Thorlindsson 1994; Bridge, Goldstein, and Brent 2006; Evans, Hawton, and Rodham 2004), even after controlling for other measures of social integration, regulation, and psychological distress (e.g., Bearman and Moody 2004; Bjarnason 1994). A few studies also demonstrate a positive association between exposure to suicidal behavior in role models and an individual's likelihood of attempting suicide (Bearman and Moody 2004). These studies add to our understanding of sociological influences on suicide, but they fail to examine who is most vulnerable to suggestion and how long effects may linger, and they are often limited by the use of cross-sectional data.

Three studies employ longitudinal data and thus shed further light on suicide suggestion within the adolescent suicide process. Brent and colleagues (1989) had the rare opportunity to collect data immediately following a suicide at a high school. Although they were unable to measure students' predispositions to suicide prior to a peer's suicide, their findings suggest that suicide suggestion can spread rapidly and then gradually lose some of its effect. More recently, Niederkrotenthaler and colleagues (2012) found that young children exposed to a parent's suicidal behavior were far more likely to develop suicidal behaviors over time than were their counterparts. This work, however, is primarily epidemiological and fails to control for potentially significant confounding factors, such as social integration. Finally, Thompson and Light (2011) examined which factors are associated with adolescent nonfatal suicide attempts and found that role models' attempts significantly increase adolescents' likelihood of attempting suicide, net of respondents' histories of suicidal thoughts and many other factors. These studies provide insights into exposure to a role model's suicidal behavior, but questions of who is most vulnerable and how long that vulnerability lasts remain open, and the role suggestion plays as an aspect of social integration remains unacknowledged.

Similarity between individuals and role models. A primary limitation in the existing literature on suicide suggestion is its failure to determine whether the similarity between friends' or family members' suicidal behaviors is due to the tendency for individuals to form friendships with people they are similar to. This proverbial "birds of a feather" is often the case for teens, who select friends and peer groups based on how similar potential friends are to themselves (Crosnoe, Frank, and Mueller 2008; Joyner and Kao 2000). Research shows that adolescent friendships tend to be homophilous in terms of depression levels (Schaefer, Kornienko, and Fox 2011) and aggression (Cairns et al. 1988). The effect of suicide suggestion on an adolescent's suicidal behaviors may thus be due to unobserved preexisting similarities between friends. To address this limitation, we focus on the development of suicidal behaviors in a sample of adolescents with no documented history of suicidality, to avoid (to the extent possible with survey data) confounding the observed effect of suicide suggestion with selection into friendships. Answering this crucial question, whether suicide suggestion contributes to the development of suicidal behaviors, is a central goal of this study.

Temporal limits. In the process of discerning how suggestion shapes adolescent suicidality, it is useful to consider whether effects of suggestion via personal role models linger as time passes, and for whom. Given past research, suggestive effects likely have temporal limitations. Previous studies on effects of media exposure generally find that spikes in suicide rates last between two and four weeks (Phillips 1974; Stack 1987). Significant others tend to have a greater impact on individuals than do nonsignificant others (Turner 2010), so it is reasonable to expect effects of personal role models will last longer than suicides publicized in the media. We thus utilize the Add Health survey to test whether the impact of a role model's suicide attempt is observable after approximately one year and six years.

Family versus friends. Generally, studies of suicide suggestion do not distinguish between effects of a family member's versus a friend's suicide attempt on those exposed. Given that past research demonstrates that "the influence of friends surpasses that of parents" by mid-adolescence (Crosnoe 2000:378), and friends' influence is strongly linked with teen delinquency, health behaviors, and pro-social behaviors (Frank et al. 2008; Giordano 2003; Haynie 2001; Mueller et al. 2010), we would expect to see differences based on an individual's relationship to the role model. It is plausible, given the extant research on adolescents and peer influence, that a friend's suicidal behavior provides a more salient model for imitating than would family. We thus analyze the two types of role models separately.

Gender differences. The final aspect deserving greater attention is potential gender differences in suggestion and suicidality. Little research emphasizes potential gender differences in how adolescents develop suicidal behaviors, despite the fact that key differences exist in suicidal behaviors between adolescent boys and girls (Baca-Garcia et al. 2008); for example, girls are more likely than boys to report nonfatal suicide attempts, whereas boys are more likely to experience fatal suicides. Another important reason to consider how suicide suggestion affects boys and girls stems from differences in boys' and girls' friendships. Girls tend to have fewer, but more intimate, emotionally laden friendships, whereas boys tend to maintain less emotional and more diffuse networks focused around shared activities (Crosnoe 2000). Moreover, girls tend to be more sensitive to others' opinions (Gilligan 1982) and are more easily influenced by peers than are boys (Maccoby 2002). These findings suggest girls may be more susceptible than boys to role models' suicide attempts.

In summary, this study shifts the sociological focus away from the protective nature of social ties toward the potential harm these ties can have on individuals. Specifically, we elaborate how exposure to suicidal behaviors shapes adolescent suicidality by identifying how, when, and for whom suicide suggestion matters. Our strategy includes (1) examining the development of suicidal behaviors in a sample of youth with no suicidal behaviors at Time I; (2) determining how long the effect of suggestion lasts; and if (3) the type of role model or (4) gender makes a difference in the process. Answers to these questions will help us understand how social relationships work in daily life to both protect and, sometimes, put individuals at risk of suicidality, thereby moving us closer to a robust sociological theory of suicide.

Methods

Data

This study employs data from Waves I, II, and III of the National Longitudinal Study of Adolescent Health (Add Health). Add Health contains a nationally representative sample of U.S. adolescents in grades 7 through 12 in 132 middle and high schools in 80 different communities. From a list of all schools containing an 11th grade in the United States, Add Health selected a nationally representative sample of schools using a school-based, cluster sampling design, with the sample stratified by region, urbanicity, school type, ethnic composition, and size.

The preliminary in-school survey collected data from all students in all Add Health high schools (*n* = 90,118 students) in 1994 to 1995; from this sample, a nationally representative subsample was interviewed at Wave I (*n* = 20,745), shortly after the in-school survey. Wave II followed in 1996 and collected information from 14,738 Wave I participants. Some groups of respondents were generally not followed up at Wave II; the largest of these were Wave I 12th graders, who had generally graduated high school by Wave II. Wave III was collected in 2001 to 2002 and followed up the Wave I in-home respondents (including respondents excluded from Wave II) who were then approximately age 18 to 23 years. Additional information about Add Health can be found in Harris and colleagues (2009).

Sample Selection

We used several sample selection filters to produce analytic samples that allow us to assess suicide suggestion in adolescence. First, we selected respondents with valid sample weights so we could properly account for the complex sampling frame of the Add Health data. Second, we used longitudinal data analysis; as such, we restricted our sample to adolescents who participated in Waves I and II of Add Health for our analyses of Wave II outcomes, and Waves I, II, and III for our analyses of Wave III outcomes. Among respondents, 10,828 had valid sample weights and participated in all three waves of Add Health. Our third selection filter selected only adolescents with no suicidal thoughts or attempts at Wave I, so the time order of events is preserved such that we can determine whether suicide suggestion plays a role above and beyond preexisting vulnerabilities to suicidality. This restriction reduced our analytic sample to 9,309 respondents. With this sample restriction, our models are not estimating the potential for role models to maintain or dissolve an adolescent's suicidal thoughts. Instead, our models estimate whether role models' behaviors at Wave I are associated with the development of previously undocumented suicidal thoughts and attempts at later waves. This also allows us to control for potential unmeasured factors that may shape both who adolescents choose as friends and their vulnerability to suicide (following the logic of classic ANCOVA; cf. Shadish, Campbell, and Cook 2002). Our final selection filter excluded adolescents missing any key independent variables.

These restrictions have the potential to bias our sample, but they also enable our analysis of critical aspects of suicidal behaviors in adolescence. To assess any potential bias, Table 1 presents descriptive statistics for the entire Wave I sample and our Wave II and Wave III analytic samples. The only substantial difference between the Wave I Add Health sample and our analytic sample is the lower incidence of suicidal thoughts and attempts at Waves II and III due to our restricting our analyses to adolescents with no suicidal thoughts at Wave I. Our analytic samples do not vary substantially from the entire Wave I sample in terms of average levels of emotional distress or demographic variables.

Measures

Dependent variables. We analyze two dependent variables: *suicidal ideation* and *suicide attempts* at Wave II and Wave III. *Suicidal ideation* is based on adolescents' responses to the question: "During the past 12 months, did you ever seriously think about committing suicide?" Adolescents who answered "yes" were coded 1 on a dichotomous outcome indicating suicidal ideation. Adolescents who reported having suicidal thoughts were then asked, "During the past 12 months, how many times did you actually attempt suicide?" Answers ranged from 0 (0 times) to four (six or more times). We recoded these responses into a dichotomous variable where 1 indicates a report of at least one suicide attempt in the past 12 months and 0 indicates no attempts. Adolescents who reported no suicidal thoughts were also coded 0 on *suicide attempts*. These variables were asked at all three waves.

Independent variables. Our first key independent variable, one of two ways we measure suicide suggestion, is *friend suicide attempt* and is based on adolescents' responses to the question: "Have any of your friends tried to kill themselves during the past 12 months?" Adolescents who responded "yes" were coded 1 on a dichotomous variable. This question was asked at all waves. For models predicting suicidal thoughts and attempts at Wave II, we rely on adolescents' responses at Wave I to preserve time order in these data. For models predicting Wave III dependent variables, we use adolescents' responses to this question at Wave II. Our second key independent measure of suicide suggestion is *family suicide attempt*. The treatment of this variable is identical to *friend suicide attempt* and is based on adolescents' responses to the question: "Have any of your family tried to kill themselves during the past 12 months?"

Our models also control for protective factors for suicide suggested by prior research. Following Durkheim's ideas about the importance of social integration as a protective factor for suicide, we measure adolescents' family integration, how close they feel to their friends, and their religious attendance. Our *family integration scale* (Cronbach's alpha = .769) is based on four items that measure how integrated adolescents are in their families (Bjarnason 1994). Adolescents were asked how much they feel their parents care about them, how much people in their family understand them, whether they have fun with their family, and whether their family pays attention to them. Responses were coded so that a higher value on the scale indicates a higher feeling of family caring. Our measure of adolescents' relationships with their friends, *friends care,* is based on adolescents' responses to the question, "How much do you feel that your friends care about you?" Higher values on this measure indicate a higher feeling of caring friends. *Religious attendance* measures how often adolescents attend religious services. Responses range from "never" to "once a week, or more." Items were coded so that a higher value on this measure indicates more frequent religious attendance.

Table 1. Weighted Descriptive Statistics for Key Variables

	Full Wave 1 Sample (mean)		Wave 2 Analytic Sample (mean)		Wave 3 Analytic Sample (mean)	
	Boys	Girls	Boys	Girls	Boys	Girls
Suicide Ideation, W1	.103	.165	.000	.000	.000	.000
Suicide Attempt, W1	.021	.057	.000	.000	.000	.000
Suicide Ideation, W2	.083	.146	.052	.092	.051	.091
Suicide Attempt, W2	.019	.051	.009	.026	.006	.026
Suicide Ideation, W3	.068	.072	.060	.060	.057	.061
Suicide Attempt, W3	.010	.025	.008	.021	.009	.020
Age, W1	15.180	15.370	15.290	15.120	15.130	15.010
	(1.610)	(1.710)	(1.620)	(1.580)	(1.530)	(1.530)
White	.667	.676	.673	.680	.667	.673
African American	.138	.151	.138	.157	.146	.162
Asian American	.044	.038	.040	.036	.039	.035
Hispanic	.118	.109	.115	.105	.115	.108
Other Race/Ethnicity	.033	.025	.034	.022	.033	.021
Parents' Education	2.867	2.853	2.880	2.867	2.935	2.910

	(1.284)	(1.261)	(1.236)	(1.239)	(1.240)	(1.248)
Lives with Two Biological Parents	.581	.571	.594	.588	.590	.602
Same-Sex Attraction, W1	.075	.048	.065	.039	.066	.038
GPA, W1	2.727	2.925	2.751	2.972	2.780	3.001
	(.798)	(.764)	(.766)	(.747)	(.753)	(.740)
Emotional Distress, W1	28.933	30.813	28.110	29.423	27.910	29.110
	(6.868)	(8.137)	(5.910)	(6.990)	(5.830)	(6.800)
N	5,042	5,694	4,301	4,523	3,855	4,075

Note: Standard deviations are in parentheses.
Source: The National Longitudinal Study of Adolescent Health.

In addition to measures of social integration, we control for several known risk factors for suicide. These include adolescents' reports of *same-sex attraction* (at Wave I) or identity as gay, lesbian, or bisexual (which was only collected at Wave III). At Wave I, adolescents were asked whether they had "ever had a romantic attraction to a female?" or "... to a male?" These questions were used to identify adolescents who experienced some form of same-sex attraction (Pearson, Muller, and Wilkinson 2007). At Wave III, adolescents were asked to choose a description that fit their sexual identity, from 100 percent homosexual to 100 percent heterosexual (with not attracted to males or females as an option). Adolescents who reported being "bisexual," "mostly homosexual (gay), but somewhat attracted to people of the opposite sex," or "100 percent homosexual (gay)" were coded 1. Heterosexual, asexual, and mostly heterosexual adolescents were coded 0.

Because emotional distress may increase an adolescent's likelihood of becoming suicidal, we control for *emotional distress* in all models. *Emotional distress* is measured by a 19-item abridged Center for Epidemiological Studies-Depression (CESD) scale (Cronbach's alpha =.873). Add Health, at Waves I and II, posed a series of questions asking respondents how often "you didn't feel like eating, your appetite was poor," "you felt that you were just as good as other people," and "you felt depressed." Positive items were reverse coded, so a higher score on every question indicates higher emotional distress. Items were then summed for adolescents who provided a valid answer to every question in the scale.

Finally, all models control for several demographic and personal characteristics, including educational attainment measures, family structure, age, race/ethnicity, and parents' education levels. Overall *grade point average* (GPA) is a self-reported measure and has the standard range of 0 to 4 (indicating the highest possible grade). An indicator for whether the adolescent successfully graduated from high school and if they attended some college is included in the models predicting suicidal behaviors at Wave III. Because of the age range of the sample, some students had not had time to complete a college degree; however, all had an opportunity to begin their college coursework and graduate from high school.

Family structure captures whether respondents lived in a two-biological-parent family, a single-parent family, a family that includes step-parents, or another family type at Wave I. *Race/ethnicity* was coded as five dichotomous variables: Latino/a, Black, Asian American, and other race or ethnicity, with White as the reference category. We took *parents' education* from the parent questionnaire and used the maximum value in the case of two parents. If this information was missing from the parent questionnaire, we used students' reports of their parents' education level. We coded parents' education as (0) never went to school; (1) less than high school graduation; (2) high school diploma or equivalent; (3) some college, but did not graduate; (4) graduated from a college or university; and (5) professional training beyond a four-year college or university.

Analytic Plan

Our goal with these analyses is to investigate whether a role model's suicide attempt is associated with the development of suicidal behaviors at Times II and III in a sample of adolescents with no suicidal behaviors at Time I. We also examine how long the increase in vulnerability lasts after exposure to a role model's suicide attempt, whether the type of role model makes a difference, and if there is variation in these processes by gender. To investigate these questions we estimate a series of nested logistic regression models with a sample of adolescents with no history of suicidal thoughts at Wave I. Because we are interested in (and anticipate based on prior literature) gender differences in what leads adolescents to contemplate suicide, we estimate all models separately by gender. As a first step, we estimate the bivariate relationships between a role model's suicide attempt (at Wave I or II) and an adolescent's likelihood of suicide ideation and attempt (at Waves II and III) to determine whether suicide suggestion is part of the process of developing suicidal behaviors over time. Next, we add a set of demographic, personal, and social characteristics to the model to determine how robust the impact of suicide suggestion is to potentially confounding risk and protective factors.[2]

Because Add Health data were collected using a complex survey design (described earlier), we estimate all models using the SAS SurveyLogistic Procedure (An 2002) to obtain appropriate estimates and standard errors (Bell et al. 2012). The survey logistic procedure is similar to traditional logistic regression, except for the handling of the variance. We estimated variance using a Taylor expansion approximation that computes variances within each stratum and pools estimates together (An 2002). This method accounts for dependencies within the data due to the complex survey design. Our models also include normalized sample weights to compensate for the substantial oversampling of certain populations. These weights render our analyses more representative of the U.S. population than would unweighted analyses that fail to correct for Add Health's oversampled populations.

Results

To begin our investigation of suicide suggestion, we first examine the roles of family members' and friends' suicide attempts in adolescent girls' and boys' suicidal behaviors at Wave II, before turning to boys' and girls' behaviors at Wave III. Among boys, reports of a new suicidal attempt were extremely rare; only 1 percent of boys reported a suicide attempt at Wave II after reporting no suicidal thoughts at Wave 1. For this reason, we focus most heavily on suicidal thoughts and examine suicide attempts only among adolescent girls. The models for boys' suicidal attempts are available from the authors by request.

Suicidal Behaviors at Wave II

Table 2 presents odds ratios from logistic regressions predicting suicide ideation and suicide attempts for girls and boys. As a first step, we estimate the bivariate relationship between family members' suicide attempts (Wave I) and adolescents' suicidal thoughts and attempts a year later (Wave II) (see Models 1, 4, and 7 in Table 2). A family member's attempted suicide (Model 1) significantly increases the likelihood that adolescent girls report suicidal thoughts at Wave II; however, it is not associated with suicide attempts at Wave II (Model 4). On average, girls who reported that a family member attempted suicide at Wave I are 2.994 times more likely to report suicidal thoughts at Wave II than are girls who did not experience a family member's suicide attempt. This pattern is not found among boys. For boys, we find no significant relationship between a family member's suicide attempt and boys' likelihood of reporting suicidal thoughts. This is our first piece of evidence for gender differences in suicide suggestion.

Next we turn to friends as role models for suicide suggestion. For girls, a friend's suicide attempt significantly increases their likelihood of reporting suicidal thoughts (Model 2) and attempts (Model 5). For boys, experiencing a friend's suicide attempt has a significant and positive relationship to boys' likelihood of reporting suicidal thoughts (Model 8). These significant bivariate relationships indicate that *who* the role model is may condition the likelihood that suicides spread through social relationships in gendered ways. Our next step is to evaluate whether these relationships maintain their significance once potential risk and protective factors are held constant in our models.

Substantively, our findings do not change after the addition of important controls.[3] On average, adolescent girls are 2.129 times more likely to report suicidal *thoughts* after experiencing a family member's attempted suicide, and 1.561 times more likely after experiencing a friend's suicide attempt, net of all other variables (Model 3). Girls' reports of suicide *attempts,* on average, are significantly related to friends' suicide attempts, but not family members' attempts, net of all other variables, confirming in Model 6 the bivariate relationships observed in Models 4 and 5. For girls, the relationship between suicide suggestion, via family or friend role models, is robust to many vital risk and protective factors for suicide.

For boys, the story is similar. The bivariate relationships observed in Models 7 and 8 are robust to the addition of control variables. Boys remain affected by a friend's suicide attempt at Wave I. Specifically, a friend's suicide attempt renders boys 1.649 times more likely to report suicidal thoughts at Wave II. The suicide attempt of a family member remains insignificant (confirming associations found in Model 7).

Overall, these findings suggest that suicide suggestion is associated with the development of suicidal behaviors within a year or so of a role model's suicide attempt, particularly when the role model is a friend. Significant gender differences do emerge: girls appear more sensitive than boys to familial role models.

Suicidal Behaviors at Wave III

In the analyses presented in Table 3, we investigate the impact a role model's suicide attempt at Wave II has on respondents' suicidal thoughts and attempts at Wave III, as respondents are entering early adulthood. These models help us understand the temporality of suicide suggestion, while also allowing us to establish a clear time order between an adolescent's history of suicidal thoughts (Wave I), the experience of a friend's or family member's suicide attempt (Wave II), and subsequent suicidal behaviors (Wave III).

Table 2. Odds Ratios from Models Predicting Suicidal Thoughts and Attempts a

	Girls					
	Suicide Ideation			Suicide Attempt		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Suicide Suggestion						
Family Suicide Attempt	2.994***		2.129**	1.069		.535
Friend Suicide Attempt		2.054***	1.561**		3.214***	2.577***
Background						
Age			.733***			.679***
African American			.625*			1.041
Asian American			.966			1.580
Latino\a			.811			1.082

Other Race or Ethnicity			.692			1.332
Parents' Education Level			.967			.865
Same-Sex Attraction			1.660			1.281
GPA			.870			.967
Social Integration						
Religious Attendance			.996			.900
Single-Parent Family			1.499[*]			1.145
Step-Parent Family			1.295			1.868
Other Family Structure			1.050			1.578
Family Integration Scale			.877			.681
Friends Care			1.204			1.216
Psychological Factors						
Emotional Distress			1.067[***]			1.067[***]

Likelihood	2708.714	2698.139	2499.105	1073.977	1039.891	947.583
Response Profile (n=1/n=0)	351/4172	351/4172	351/4172	100/4423	100/4423	100/4423
N	4,523	4,523	4,523	4,523	4,523	4,523

Note: All independent variables measured at Wave I.
Source: The National Longitudinal Study of Adolescent Health.

**p* <.05;

***p* <.01;

****p* <.001 (two-tailed tests).

Table 3. Odds Ratios from Models Predicting Suicidal Thoughts and Attempts am

	Girls						Bo
	Suicide Ideation			Suicide Attempt			Su
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	M
Suicide Suggestion							
Family Suicide Attempt	.725		.466	1.298		.779	1.
Friend Suicide Attempt		1.978***	1.546		1.794	1.254	
Background							
Age			.811***			.824	
African			.535*			.693	

Asian American			1.286			4.808***	
Latino\a			.804			.698	
Other Race or Ethnicity			.678			<.001***	
Parents' Education Level			1.220*			1.173	
Gay, Lesbian, Bisexual Identity (W3)			2.879**			2.917	
GPA			.840			.645	
High School Dropout (W3)			1.557			2.688*	
Some College (W3)			1.063			1.264	
Social Integration							
Religious Attendance			.845			.883	
Single-Parent Family			1.200			2.796*	
Step-Parent Family			.995			2.560*	
Other							

Family							
Other Family Structure			1.447			1.939	
Not Currently Married or Cohabiting (W3)			1.309			1.173	
Family Integration Scale			.871			1.015	
Friends Care			1.014			1.481	
Psychological Factors							
Emotional Distress			1.041**			1.026	
–2 Log Likelihood	1841.515	1821.903	1709.455	794.630	789.706	709.774	16
Response Profile (n=1/n=0)	202/3873	202/3873	202/3873	59/4016	59/4016	59/4016	19
N	4,075	4,075	4,075	4,075	4,075	4,075	3,

Note: All independent variables measured at Wave II unless otherwise noted.
Source: The National Longitudinal Study of Adolescent Health.

*p <.05;

**p <.01;

***p <.001 (two-tailed tests).

friend's suicide attempt at Wave II and girls' reports of suicidal thoughts at Wave III. This finding, however, does not hold in full models, although the odds ratio is in the expected direction (OR = 1.546) and the *p*-value is very close to the threshold for statistical significance (p >.055) (Model 3 in Table 3). We further investigated the change in statistical significance between the bivariate and saturated models in analyses not presented here (but available from the authors by request). We found that adolescent girls' emotional distress at Wave II explains the impact of a friend's suicide attempt on girls' likelihood of reporting suicidal thoughts at Wave III, net of other key controls. The significant effect of a friend's suicide attempt on girls' likelihood of suicidal thoughts remains until emotional distress is included in the model. This suggests that emotional distress may serve as an important mechanism through which suicide suggestion operates, particularly for girls.

Our models from Wave III suggest that the increased risk of suicide suggestion found over the short run (in Table 2) fades with time. Six years later, we find little evidence that experiencing a role model's suicide attempt, whether friend or family member, has a long-term effect, except perhaps for girls for whom it is mediated by emotional distress.

Discussion

Within the sociology of suicide, social integration and regulation are often emphasized as the primary social forces that protect or put individuals at risk of suicide. These Durkheimian mechanisms are undoubtedly important (Bearman 1991; Pescosolido 1990; Pescosolido and Georgianna 1989; Wray et al. 2011), but much research on the spread of health behaviors implicates social ties as not just mechanisms for social support, but also potential conduits for the spread of suicidal behaviors via suicide suggestion, illuminating another side to social integration. We find that suicide attempts of role models—primarily friends— are in fact associated with adolescents' development of suicidal thoughts and, in some cases, attempts. Effects of suicide suggestion appear to fade with time, girls are more vulnerable to suicide suggestion than boys, and the type of role model—for teenagers at least—matters. Our findings suggest that social relationships, contra Durkheim, are not always protective against suicide, at least not when significant others exhibit suicidal tendencies. This reinforces the idea that the structure—and content—of social networks conditions their role in preventing suicidality. Specifically, social ties can be conduits of not just social support but also antisocial behaviors, like suicidality.

Our study has four primary implications for advancing the sociological understanding of suicide. Our most essential contribution to the literature on suicide suggestion via personal role models is the evidence we provide indicating that being aware of a role model's suicide attempt is associated with the development of suicidal thoughts and sometimes attempts. This relationship is robust to many measures of risk and protective factors. Experiencing the suicide attempt of a significant other may serve as a vehicle for learning a way to deal with distressing life events—by becoming suicidal (Jamison 1999). Future research should continue to probe the question of how suicide suggestion contributes to the development of suicidality. Many potential mechanisms—social learning, imitation, and emotional contagion—may underlay the observed association between role models and those exposed to their suicidality. Qualitative research, in particular, may provide valuable insights into which potential mechanisms promote the spread of suicidality via social ties. Understanding how and when suicide suggestion becomes salient to youths' suicidality would greatly help practitioners prevent suicides. Our study provides a first step toward this larger goal.

In addition to providing insights into suicide suggestion as an important mechanism in the adolescent suicide process, our study has implications for understanding the temporality of suicide suggestion via individuals' role models. Previous research on suicide rates and media exposure found effects of suicide suggestion tend to last two to four weeks (Phillips 1974; Stack 1987). Considering the potential differences in connectedness derived from face-to-face relationships and direct contact versus mediated sources, we hypothesized that personal role models would have a stronger, or longer lasting, effect on adolescents exposed to their behavior. In fact, our findings suggest that having a friend attempt suicide has a longer lasting effect than reading about a suicide in the paper or seeing a fictive suicide on television. We find that effects of a friend's or family member's suicide attempt last at least one year, if not more—considerably longer than the effect of exposure via the media documented in prior research. By six years, however, the effect of a friend's or family member's suicide attempt appears to fade in significance. Among adolescent girls, however, a friend's suicide attempt may continue to shape suicidal thoughts even six years later; notably, this effect is explained by girls' emotional distress levels. Future research should examine this pattern in more detail, as this finding suggests an indirect, but potentially important, long-term impact of suicide suggestion via girls' emotional distress.

Perhaps it is not shocking that we do not find strong evidence that effects of role models' suicide attempts last over the long run. Teens who survive the first year (or so) following a friend's suicide attempt may be, or become, emotionally resilient. By early adulthood, a role model's suicide attempt in adolescence may no longer be central to one's daily life, a life no longer constrained within the bounds of high school. Research on contagion generally focuses on relatively bounded social spaces—like Native American reservations, mental wards, or high schools—and finds these spaces are at higher risk of geographic-temporal suicide clustering (e.g., Gould et al. 1990). Outside of relatively bounded social environments, do effects of role models' suicides spread via social ties? Investigating the role of exposure to suicides inside and out of bounded social contexts would add more depth to our understanding of how suicides—and potentially other behaviors—become socially contagious.

Our third major contribution to the literature comes from our emphasis on the role of gender in the suicide suggestion process. Given that boys and girls experience peer relationships differently (Crosnoe 2000), understanding how a social mechanism, such as suicide suggestion, differs for boys and girls is crucial to arriving at a full understanding of the development of adolescent suicidality. In fact, we find significant gender differences in the role of suicide suggestion: suggestion appears more salient to girls. Among boys, friends are the only relevant personal role models for triggering the development of suicidal thoughts; girls' suicidal behaviors, on the other hand, are influenced by both family and friends. Moreover, among girls, suicidal thoughts *and attempts* are associated with suicide suggestion. Finally, effects of a friend's or family member's suicide attempt may last longer for girls.

Although we found girls were more vulnerable, absent an observed history of suicidal thoughts, boys were not immune to suicide suggestion. Note that Thompson and Light (2011), who analyzed suicide attempts net of prior suicidal thoughts, found that boys and girls responded similarly to a role model's suicide attempt. This suggests the role of gender may change at different points in the suicidal process and that a predisposition toward suicidality may be particularly important for understanding those differences.

Why would girls be more vulnerable than boys to suicide suggestion? A definitive answer to this question is beyond the scope of this article, but we can suggest some theoretical considerations that may help explain this variation and offer paths for future research. Because girls develop and maintain more intense intimate relationships (Crosnoe 2000), they may be more primed to "take the role of the other" and hence may be more vulnerable to suggestive mechanisms, including developing emotional distress that sustains the original suggestive triggers. For boys, having relationships that are far less emotionally anchored may reduce or mitigate the effects of suggestion, which raises vital questions about which mechanisms are more salient in the development of boys' suicidal thoughts. Future research should continue to examine the complex role gender plays in the adolescent suicide process, as this may help determine different strategies for preventing suicides.

Our fourth and final major contribution to the sociology of suicide stems from our examination of how different role models— friends and family members—vary in terms of their importance in the suicide suggestion process. Our findings indicate that peers may be more meaningful than family to adolescents, for both boys and girls. Social psychology has long shown that behavior is more strongly shaped by members of reference groups central to the formation and maintenance of one's identity (Stryker 1980). To be sure, a teen's family consists of similar individuals whom the teen may identify with, but research on adolescents clearly demonstrates that purposive efforts to differentiate oneself from one's family are accompanied by concomitant identification with peers. This is not to say that a family member's suicidal tendencies are not distressing in adolescence. For example, we find that for adolescent girls, over the short run, a family member's suicide attempt increases their likelihood of reporting suicidal thoughts (but not attempts) one year later. Yet taken as a whole, our findings indicate that friends' suicide attempts are more influential than family members' suicide attempts in adolescents' lives, at least once adolescents' Wave I suicidality is controlled.

Limitations

Although our findings provide new and important insights into the sociology of suicide, this study is not without its limitations. First, and perhaps most obvious, we are limited to analyzing respondents' suicidal behaviors because we have no information on Add Health respondents who commit suicide. Individuals who report suicidal thoughts or have a history of nonfatal suicide attempts are significantly more likely to commit suicide, but fatal suicide attempts are most common among individuals with no history of nonfatal suicide attempts. Generalizing these findings to the spread of suicide deaths should thus be done with caution. Furthermore, there is attrition in the Add Health sample between waves, and given the higher completion rate among male suicide attempters, more boys than girls may be missing from our analyses due to a completed suicide. Additionally, respondents who actually commit suicide may have been the most likely to be affected by suicide suggestion. Unfortunately, we could find no information from Add Health on whether suicide, or even death, played a significant role in sample attrition. Fortunately, the rarity of suicide among adolescents reduces the risk of this substantially biasing our findings. However, this discussion highlights the significance of finding a way to compare the "lethality" of all types of role models, from the personal to the media-based. Future data collection efforts should note this key gap in the literature.

Our second limitation is related. We chose to focus on friends' and family members' suicide attempts, rather than actual suicides, for practical reasons. Very few respondents reported having a friend or family member complete suicide. This fact may affect our findings on the importance of suicide suggestion. The power of suicide suggestion in the case of a suicide may be greater than the power of suggestion based on a nonfatal suicide attempt. If anything, our findings may thus underrepresent the potential salience of suicide suggestion as a social mechanism in suicidal behaviors.

Finally, although we did our best to account for adolescents' vulnerability to suicide, we are limited by available data. Specifically, we analyzed a sample of adolescents who reported no suicidal thoughts at Wave I in an attempt to parse out effects of selection into friendships from the influence those friendships may have on an individual. Some adolescents with a history of suicidality, perhaps prior to Wave I, may have been included in our sample. Our study provides one of the best efforts to date to isolate selection from the effect of suicide suggestion, but further investigation of these issues is needed before we can be confident that suggestion affects the development of suicidality.

Conclusions

Sociologists commonly turn to Durkheimian measures of social integration and regulation when searching for sociological explanations for suicide, but our findings indicate that suicides, like other behaviors, can spread through social relationships via suicide suggestion. Friends' and family members' suicide attempts may trigger the development of suicidal behaviors, suggesting that exposure to role models is a powerful way that drastic and deviant behaviors, like suicide, become normalized. Notably, the relationship to the role model conditions the experience of suicide suggestion. Furthermore, adolescent girls appear more susceptible than boys to adopting the suicidal behaviors they observe through social relationships. This study provides important information for the evolution of the sociology of suicide, but our findings also have vital policy implications for public health officials attempting to prevent adolescent suicide. Namely, policies and practitioners need to be sensitive to the importance of suicide attempts (and not simply suicides), particularly among peers and for girls. Additionally, the increased risk of suicidality associated with friends' suicide attempts may last a year or more, which is longer than previously thought.

For adolescents, ties do bind, but whether these ties integrate adolescents into society, with positive repercussions for their emotional well-being, or whether they promote feelings of alienation, depends in part on the qualities embedded in those ties. On the surface, these findings may appear to contradict Durkheim's sociology, given his focus on solidarity through collective effervescence. Yet, Durkheim argued that solidarity was a product of a shared, collective conscience that spreads through ritualized, emotion-laden interaction. Why should we expect deviant behavior like suicide to be precluded from the types of norms that can spread across actors? Instead, we posit that for a full understanding of how social integration works in individuals' lives to shape their life chances, we must consider not only the social support social ties provide, but also the emotions, behaviors, and values that inhere in those social relations.

Acknowledgments

Seth Abrutyn and Anna Mueller contributed equally to this work. This article is a revision of a paper presented at the 2012 annual meetings of the American Sociological Association. The authors would like to thank Marty Levin, Chandra Muller, Ken Frank, Sarah Blanchard, and six anonymous reviewers for their insightful comments and suggestions. The authors acknowledge the helpful research assistance of Cynthia Stockton.

Data and Funding

This research uses data from Add Health (http://www.cpc.unc.edu/addhealth), a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. No direct support was received from grant P01-HD31921 for this analysis. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies.

Notes

1. We are particularly grateful to an anonymous reviewer for suggesting this formulation.

2. The SAS programs used to recode and analyze all data are available from the authors by request.

3. Tables presenting odds ratios and confidence intervals are available from the authors by request.

References

An, Anthony. 2002. “Performing Logistic Regression on Survey Data with the New SURVEYLOGISTIC Procedure.” Paper 258-27, pp. 1–9. Proceedings of the 27th Annual SAS Users Group International Conference (SUGI 27), Orlando, FL, April 14–17.

Baca-Garcia, Enrique, M. Mercedes Perez-Rodriguez, J. John Mann, and Maria A. Oquendo. 2008. “Suicidal Behavior in Young Women.” *Psychiatric Clinics of North America* 31:317–31.

Baller, Robert D. and Kelly K. Richardson. 2002. “Social Integration, Imitation, and the Geographic Patterning of Suicide.” *American Sociological Review* 67:873–88.

Baller, Robert D. and Kelly K. Richardson. 2009. “The ‘Dark Side’ of the Strength of Weak Ties: The Diffusion of Suicidal Thoughts.” *Journal of Health and Social Behavior* 50:261–76.

Bearman, Peter S. 1991. “The Social Structure of Suicide.” *Sociological Forum* 6:501–524.

Bearman, Peter S. and James Moody. 2004. “Suicide and Friendships among American Adolescents.” *American Journal of Public Health* 94:89–95.

Bell, Bethany A., Anthony J. Onwuegbuzie, John M. Ferron, Qun G. Jiao, Susan T. Hibbard, and Jeffrey D. Kromrey. 2012. “Use of Design Effects and Sample Weights in Complex Health Survey Data: A Review of Published Articles Using Data From 3 Commonly Used Adolescent Health Surveys.” *American Journal of Public Health* 102:1399–1405.

Bjarnason, Thoroddur. 1994. “The Influence of Social Support, Suggestion and Depression on Suicidal Behavior among Icelandic Youth.” *Acta Sociologica* 37:195–206.

Bjarnason, Thoroddur and Thorolfur Thorlindsson. 1994. “Manifest Predictors of Past Suicide Attempts in a Population of Icelandic Adolescents.” *Suicide and Life Threatening Behavior* 24:350–58.

Brent, David A., Mary M. Kerr, Charles Goldstein, James Bozigar, Marty Wartella, and Marjorie J. Allan. 1989. “An Outbreak of Suicide and Suicidal Behavior in a High School.” *American Academy of Child and Adolescent Psychiatry* 28:918–24.

Bridge, Jeffrey A., Tina R. Goldstein, and David A. Brent. 2006. “Adolescent Suicide and Suicidal Behavior.” *Journal of Child Psychology and Psychiatry* 47:372–94.

Cairns, Robert B., Beverly D. Cairns, Holly J. Neckerman, Scott D. Gest, and Jean-Louis

Gariepy. 1988. “Social Networks and Aggressive Behavior: Peer Support or Peer Rejection?” *Developmental Psychology* 61:157–68.

Crosnoe, Robert. 2000. “Friendships in Childhood and Adolescence: The Life Course and New Directions.” *Social Psychology Quarterly* 63:377–91.

Crosnoe, Robert, Kenneth Frank, and Anna Strassmann Mueller. 2008. “Gender, Body Size, and Social Relations in American High Schools.” *Social Forces* 86:1189–1216.

Crosnoe, Robert and Monica Kirkpatrick Johnson. 2011. “Research on Adolescence in the Twenty-First Century.” *Annual Review of Sociology* 37:479–60.

Durkheim, Emile. [1897]1951. *Suicide: A Study in Sociology*. Glencoe, IL: Free Press.

Evans, Emma, Keith Hawton, and Karen Rodham. 2004. “Factors Associated with Suicidal Phenomena in Adolescents: A Systematic Review of Population-Based Studies.” *Clinical Psychology Review* 24:957–79.

Frank, Kenneth, Chandra Muller, Catherine RiegleCrumb, Anna Strassmann Mueller, and Jennifer Pearson. 2008. “The Social Dynamics of Mathematics Coursetaking in High Schools.” *American Journal of Sociology* 113:1645–96.

Gilligan, Carol. 1982. *In A Different Voice: Psychological Theory and Women’s Development*. Cambridge, MA: Harvard Press.

Giordano, Peggy C. 2003. “Relationships in Adolescence.” *Annual Review of Sociology* 29:252–81.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Gould, Madelyn S. 2001. “Suicide and the Media.” Pp. 200–224 in *Suicide Prevention: Clinical and Scientific Aspects,* edited by H. Hendin and J. J. Mann. New York: New York Academy of Science.

Gould, Madelyn S., Sylvan Wallenstein, and Marjorie Kleinman. 1990. “Time-Space Clustering of Teenage Suicide.” *American Journal of Epidemiology* 131:71–78.

Harris, Kathleen M., C. T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J. Richard Udry. 2009. “The National Longitudinal Study of Adolescent Health: Research Design.” (http://www.cpc.unc.edu/projects/ addhealth/design).

Haynie, Dana L. 2001. “Delinquent Peers Revisited: Does Network Structure Matter?” *American Journal of Sociology* 106:1013–57.

Jamison, Kay Redfield. 1999. *Night Falls Fast: Understanding Suicide*. New York: Vintage

Books.

Joyner, Kara and Grace Kao. 2000. "School Racial Composition and Adolescent Racial Homophily." *Social Science Quarterly* 81:810–25.

Lawler, Edward J. 2006. "The Affect Theory of Social Exchange." Pp. 248–67 in *Contemporary Social Psychological Theories,* edited by P. J. Burke. Stanford, CA: Stanford University Press.

Liu, Ruth X. 2006. "Vulnerability to Friends' Suicide Influence: The Moderating Effects of Gender and Adolescent Depression." *Journal of Youth and Adolescence* 35:479–89.

Maccoby, Eleanor E. 2002. "Gender and Group Process: A Developmental Perspective." *Current Directions in Psychological Science* 11:54–58.

Maimon, David and Danielle C. Kuhl. 2008. "Social Control and Youth Suicidality: Situating Durkheim's Ideas in a Multilevel Framework." *American Sociological Review* 73:921–43.

Mueller, Anna S., Jennifer Person, Chandra Muller, Kenneth Frank, and Alan Turner. 2010. "Sizing Up Peers: Adolescent Girls' Weight Control and Social Comparison in the School Context." *Journal of Health and Social Behavior* 51:64–78.

National Institute of Mental Health (NIMH). 2003. *In Harm's Way: Suicide in America.* U.S. Department of Mental and Human Services. NIH Publication no. 03-4594.

Niederkrotenthaler, Thomas, Brigitta Roderus, Kristina Alexanderson, Finn Rasmussen, and Ellenor Mittendorfer-Rutz. 2012. "Exposure to Parental Mortality and Markers of Morbidity, and the Risks of Attempted and Completed Suicide in Offspring: An Analysis of Sensitive Life Periods." *Journal of Epidemiology and Community Health* 66:232–39.

Niederkrotenthaler, Thomas and Gernot Sonneck. 2007. "Assessing the Impact of Media Guidelines for Reporting Suicides in Austria: Interrupted Time Series Analysis." *Australian and New Zealand Journal of Psychiatry* 41:419–28.

Pearson, Jennifer, Chandra Muller, and Lindsey Wilkinson. 2007. "Adolescent Same-Sex Attraction and Academic Outcomes: The Role of School Attachment and Engagement." *Social Problems* 54:523–42.

Pescosolido, Bernice A. 1990. "The Social Context of Religious Integration and Suicide: Pursuing Network Explanation." *Sociological Quarterly* 31:337–57.

Pescosolido, Bernice and Sharon Georgianna. 1989. "Durkheim, Suicide, and Religion: Toward a Network Theory of Suicide." *American Sociological Review* 54:33–48.

Phillips, David P. 1974. “The Influence of Suggestion on Suicide: Substantive and Theoretical Implications of the Werther Effect.” *American Sociological Review* 39:340–54.

Schaefer, David R., Olga Kornienko, and Andrew M. Fox. 2011. “Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily.” *American Sociological Review* 76:764–85.

Schmidtke, Armin and H. Hafner. 1988. “The Werther Effect after Television Films: New Evidence for an Old Hypothesis.” *Psychological Medicine* 18:665–76.

Shadish, William R., Donald T. Campbell, and Thomas D. Cook. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.

Stack, Steven. 1987. “Celebrities and Suicide: A Taxonomy and Analysis, 1948–1983.” *American Sociological Review* 52:401–412.

Stack, Steven. 2003. “Media Coverage as a Risk Factor in Suicide.” *Journal of Epidemiology and Community Health* 57:238–40.

Stack, Steven. 2005. “Suicide in the Media: A Quantitative Review of Studies Based on Nonfictional Stories.” *Suicide and Life Threatening Behavior* 35:121–33.

Stack, Steven. 2009. “Copycat Effects on Fictional Suicide: A Meta-Analysis.” Pp. 231–44 in *Suicide and the Creative Arts,* edited by S. Stack and D. Lester. New York: Nova Science Publishers.

Stryker, Sheldon. 1980. *Symbolic Interactionism: A Social Structural Version*. Menlo Park, CA: The Benjamin Cummings Publishing Company.

Suicide Prevention Resource Center (SPRC). 2013. *Reporting on Suicide: Recommendations for the Media*. Washington, DC: Suicide Prevention Resource Center. Retrieved April 24, 2013 (http://www.sprc.org/sites/sprc.org/files/library/sreporting.pdf).

Thompson, Martie P. and Laney S. Light. 2011. “Examining Gender Differences in Risk Factors for Suicide Attempts Made 1 and 7 Years Later in a Nationally Representative Sample.” *Journal of Adolescent Health* 48:391–97.

Thorlindsson, Thorolfur and Thoroddur Bjarnason. 1998. “Modeling Durkheim on the Micro Level: A Study of Youth Suicidality.” *American Sociological Review* 63:94–110.

Turner, Jonathan H. 2010. *Theoretical Principles of Sociology*. Vol. 2, *Microdynamics*. New York: Springer.

Wray, Mary, Cynthia Colen, and Bernice Pescosolido. 2011. “The Sociology of Suicide.” *Annual Review of Sociology* 37:505–528.

Seth Abrutyn is Assistant Professor of Sociology at the University of Memphis. As a general sociological theorist, he has long been interested in macrosociology and institutions, which has culminated in a recently published book, *Revisiting Institutionalism in Sociology*. Recently, his research interests have moved toward the sociology of suicide, including examining the processes by which suicides can spread as well as how these processes relate to and expand Durkheim's classic thesis.

Anna S. Mueller is Assistant Professor of Sociology at the University of Memphis. Her research examines how peers shape adolescent health and well-being over the transition to adulthood, with a focus on weight-control behaviors, body weight, and suicide. Her research emphasizes why and how behaviors and values spread between individuals generally using insights from social psychology. She recently published a study, with Kenneth A. Frank and Chandra Muller (*American Journal of Sociology* 2013), that investigates how schools shape adolescent friendship formation in ways that have implications for adolescent status hierarchies.

[a]The University of Memphis

Corresponding Authors:

Seth Abrutyn and Anna Mueller, Department of Sociology, Clement Hall 231, The University of Memphis, Memphis, TN 38152-3530

E-mail: seth.abrutyn@memphis.edu; anna.mueller@memphis.edu

Appendix C Table of Random Numbers

Line/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	10480	15011	01536	02011	81647	91646	69179	14194	6259
2	22368	46573	25595	85393	30995	89198	27982	53402	9396
3	24130	48360	22527	97265	76393	64809	15179	24830	4934
4	42167	93093	06243	61680	07856	16376	39440	53537	7134
5	37570	39975	81837	16656	06121	91782	60468	81305	4968
6	77921	06907	11008	42751	27756	53498	18602	70659	9065
7	99562	72905	56420	69994	98872	31016	71194	18738	4401
8	96301	91977	05463	07972	18876	20922	94595	56869	6901
9	89579	14342	63661	10281	17453	18103	57740	84378	2533
10	85475	36857	43342	53988	53060	59533	38867	62300	0815
11	28918	69578	88231	33276	70997	79936	56865	05859	9010
12	63553	40961	48235	03427	49626	69445	18663	72695	5218
13	09429	93969	52636	92737	88974	33488	36320	17617	3001
14	10365	61129	87529	85689	48237	52267	67689	93394	0151
15	07119	97336	71048	08178	77233	13916	47564	81056	9773
16	51085	12765	51821	51259	77452	16308	60756	92144	4944
17	02368	21382	52404	60268	89368	19885	55322	44819	0118

18	01011	54092	33362	94904	31273	04146	18594	29852	7158
19	52162	53916	46369	58586	23216	14513	83149	98736	2349
20	07056	97628	33787	09998	42698	06691	76988	13602	5185
21	48663	91245	85828	14346	09172	30168	90229	04734	5919
22	54164	58492	22421	74103	47070	25306	76468	26384	5815
23	32639	32363	05597	24200	13363	38005	94342	28728	3580
24	29334	27001	87637	87308	58731	00256	45834	15398	4655
25	02488	33062	28834	07351	19731	92420	60952	61280	5000
26	81525	72295	04839	96423	24878	82651	66566	14778	7679
27	29676	20591	68086	26432	46901	20849	89768	81536	8664
28	00742	57392	39064	66432	84673	40027	32832	61362	9894
29	05366	04213	25669	26422	44407	44048	37937	63904	4576
30	91921	26418	64117	94305	26766	25940	39972	22209	7150
31	00582	04711	87917	77341	42206	35126	74087	99547	8181
32	00725	69884	62797	56170	86324	88072	76222	36086	8463
33	69011	65797	95876	55293	18988	27354	26575	08625	4080
34	25976	57948	29888	88604	67917	48708	18912	82271	6542
35	09763	83473	73577	12908	30883	18317	28290	35797	0599
36	91567	42595	27958	30134	04024	86385	29880	99730	5553
37	17955	56349	90999	49127	20044	59931	06115	20542	1805
38	46503	18584	18845	49618	02304	51038	20655	58727	2816

39	92157	89634	94824	78171	84610	82834	09922	25417	4413
40	14577	62765	35605	81263	39667	47358	56873	56307	6160
41	98427	07523	33362	64270	01638	92477	66969	98420	0488
42	34914	63976	88720	82765	34476	17032	87589	40836	3242
43	70060	28277	39475	46473	23219	53416	94970	25832	6997
44	53976	54914	06990	67245	68350	82948	11398	42878	8028
45	76072	29515	40980	07391	58745	25774	22987	80059	3991
46	90725	52210	83974	29992	65831	38857	50490	83765	5565
47	64364	67412	33339	31926	14883	24413	59744	92351	9747
48	08962	00358	31662	25388	61642	34072	81249	35648	5689
49	95012	68379	93526	70765	10593	04542	76463	54328	0234
50	15664	10493	20492	38391	91132	21999	59516	81652	2719
51	16408	81899	04153	53381	79401	21438	83035	92350	3669
52	18629	81953	05520	91962	04739	13092	97662	24822	9473
53	73115	35101	47498	87637	99016	71060	88824	71013	1873
54	57491	16703	23167	49323	45021	33132	12544	41035	8078
55	30405	83946	23792	14422	15059	45799	22716	19792	0998
56	16631	35006	85900	98275	32388	52390	16815	69298	8273
57	96773	20206	42559	78985	05300	22164	24369	54224	3508
58	38935	64202	14349	82674	66523	44133	00697	35552	3597

59	31624	76384	17403	53363	44167	64486	64758	75366	7655
60	78919	19474	23632	27889	47914	02584	37680	20801	7215
61	03931	33309	57047	74211	63445	17361	62825	39908	0560
62	74426	33278	43972	10119	89917	15665	52872	73823	7314
63	09066	00903	20795	95452	92648	45454	09552	88815	1655
64	42238	12426	87025	14267	20979	04508	64535	31355	8606
65	16153	08002	26504	41744	81959	65642	74240	56302	0003
66	21457	40742	29820	96783	29400	21840	15035	34537	3331
67	21581	57802	02050	89728	17937	37621	47075	42080	9740
68	55612	78095	83197	33732	05810	24813	86902	60397	1648
69	44657	66999	99324	51281	84463	60563	79312	93454	6887
70	91340	84979	46949	81973	37949	61023	43997	15263	8064
71	91227	21199	31935	27022	84067	05462	35216	14486	2989
72	50001	38140	66321	19924	72163	09538	12151	06878	9190
73	65390	05224	72958	28609	81406	39147	25549	48542	4262
74	27504	96131	83944	41575	10573	08619	64482	73923	3615
75	37169	94851	39117	89632	00959	16487	65536	49071	3978
76	11508	70225	51111	38351	19444	66499	71945	05422	1344
77	37449	30362	06694	54690	04052	53115	62757	95348	7866
78	46515	70331	85922	38329	57015	15765	97161	17869	4534
79	30986	81223	42416	58353	21532	30502	32305	86482	0517

80	63798	64995	46583	09765	44160	78128	83991	42865	9252
81	82486	84846	99254	67632	43218	50076	21361	64816	5120
82	21885	32906	92431	09060	64297	51674	64126	62570	2612
83	60336	98782	07408	53458	13564	59089	26445	29789	8520
84	43937	46891	24010	25560	86355	33941	25786	54990	7189
85	97656	63175	89303	16275	07100	92063	21942	18611	4734
86	03299	01221	05418	38982	55758	92237	26759	86367	2121
87	79626	06486	03574	17668	07785	76020	79924	25651	8332
88	85636	68335	47539	03129	65651	11977	02510	26113	9944
89	18039	14367	61337	06177	12143	46609	32989	74014	6470
90	08362	15656	60627	36478	65648	16764	53412	09013	0783
91	79556	29068	04142	16268	15387	12856	66227	38358	2247
92	92608	82674	27072	32534	17075	27698	98204	63863	1195
93	23982	25835	40055	67006	12293	02753	14827	22235	3507
94	09915	96306	05908	97901	28395	14186	00821	80703	7042
95	50937	33300	26695	62247	69927	76123	50842	43834	8665
96	42488	78077	69882	61657	34136	79180	97526	43092	0409
97	46764	86273	63003	93017	31204	36692	40202	35275	5730
98	03237	45430	55417	63282	90816	17349	88298	90183	3660
99	86591	81482	52667	61583	14972	90053	89534	76036	4919

100	38534	01715	94964	87288	65680	43772	39560	12918	8653

Glossary

Abductive analysis

A qualitative data analysis approach in which researchers produce theoretical hunches for unexpected research findings and develop them with a systematic analysis of variation across a study in relation to a broad range of theories and interchange with other scholars.

Adaptive research design

A research design that develops as the research progresses.

Alternate-forms reliability

A procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

Anomalous findings

Unexpected patterns in data.

Anonymity

A provision of research, in which no identifying information is recorded that could be used to link respondents to their responses.

Applied research

Research conducted using the scientific method and addressing immediate, practical concerns, such as determining whether one program or policy has a more desirable impact than another.

Assignment variable

The variable used to specify a cutoff score for eligibility for some treatment in a regression–discontinuity design.

Association

A criterion for establishing a nomothetic causal relationship between two variables: Variation in one variable is related to variation in another variable.

Authenticity

When the understanding of a social process or social setting is one that reflects fairly the various perspectives of participants in that setting.

Availability sampling

A nonprobability sampling method in which elements are selected on the basis of convenience.

Back matter

The section of an applied research report that may include appendixes, tables, figures, and the research instrument(s).

Balanced response choices

When a fixed-choice survey question has an equal number of responses to express positive and negative choices in comparable language.

Bar chart

A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces.

Base number (N)

The total number of cases in a distribution.

Basic science

Research conducted using the scientific method and having the goal of figuring out what the world is like and why it works as it does.

Before-and-after design

A quasi-experimental design consisting of several before-after comparisons involving the same variables but no comparison group.

Behavior coding

An observation in which the researcher categorizes, according to strict rules, the number of times certain behaviors occur.

Belmont Report

Guidelines developed by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979 for the protection of human subjects.

Beneficence

The ethical requirement of minimizing possible harms and maximizing benefits in research involving human subjects that was included in the *Belmont Report*.

Big Data

Massive data sets accessible in computer-readable form that are produced by people, available to social scientists, and manageable with today's computers.

Bimodal

A distribution that has two nonadjacent categories with about the same number of cases, and these categories have more cases than any others.

Bipolar response options

Response choices to a survey question that include a middle category and parallel responses with positive and negative valence (can be labeled or unlabeled).

Black box evaluation

The type of evaluation that occurs when an evaluation of program outcomes ignores, and does not identify, the process by which the program produced the effect.

Case study

A setting or group that the analyst treats as an integrated social unit that must be studied holistically and in its particularity.

Case-oriented research

Research that focuses attention on the nation or other unit as a whole.

Case-oriented understanding

An understanding of social processes in a group, formal organization, community, or other collectivity that reflects accurately the standpoint of participants.

Causal effect (idiographic perspective)

When a series of concrete events, thoughts, or actions results in a particular event or individual outcome.

Causal effect (nomothetic perspective)

When variation in one phenomenon, an independent variable, leads to or results, on average, in variation in another phenomenon, the dependent variable.

Causal validity (internal validity)

Exists when a conclusion that A leads to or results in B is correct.

Census

Research in which information is obtained through responses from or information about all available members of an entire population.

Central tendency

The most common value (for variables measured at the nominal level) or the value around which cases tend to center (for a quantitative variable).

Certificate of Confidentiality

A certificate issued to a researcher by the National Institutes of Health that ensures the right to protect from legal subpoenas information obtained about high-risk populations or behaviors—except child abuse or neglect.

Ceteris paribus

Latin phrase meaning “other things being equal.”

Chi-square

An inferential statistic used to test hypotheses about relationships between two or more variables in a cross-tabulation.

Closed-ended (fixed-choice) question

A survey question that provides preformatted response choices for the respondent to circle or check.

Cluster

A naturally occurring, mixed aggregate of elements of the population.

Cluster sampling

A sampling method in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Code of Ethics

The professional code of the American Sociological Association for the treatment of human subjects by members, employees, and students; designed to comply with federal policy and revised in 1997.

Coding

The process of assigning a unique numerical code to each response to survey questions.

Cognitive interview

A technique for evaluating questions in which researchers ask people test questions and then probe with follow-up questions to learn how they understood the question and what their answers mean.

Cohort

Individuals or groups with a common starting point. Examples include college class of 1997, people who graduated from high school in the 1980s, General Motors employees who started work between the years 1990 and 2000, and people who were born in the late 1940s or the 1950s (the baby boom generation).

Cohort study

See event-based design.

Combined frequency display

A table that presents together the distributions for a set of conceptually similar variables having the same response categories; common headings are used for the responses.

Common Rule

See Federal Policy for the Protection of Human Subjects.

Comparative historical research

Research comparing data from more than one time period in more than one nation.

Comparison group

In an experiment, a group that has been exposed to a different treatment (or value of the independent variable) than the experimental group.

Compensatory rivalry (John Henry effect)

A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied some advantages and increase their efforts by way of compensation.

Complete (or covert) participant

A role in field research in which the researcher does not reveal his or her identity as a researcher to those who are observed while participating.

Complete (or overt) observer

A role in participant observation in which the researcher does not participate in group activities and is publicly defined as a researcher.

Complex mixed-methods design

Qualitative and quantitative methods are combined in a research project that uses more than one of the four basic types of mixed-methods designs or that repeats at least one of those basic types.

Compressed frequency display

A table that presents cross-classification data efficiently by eliminating unnecessary percentages, such as the percentage corresponding to the second value of a dichotomous variable.

Computer-assisted personal interview (CAPI)

A personal interview in which a laptop computer is used to display interview questions and to process responses that the interviewer types in, as well as to check that these responses fall within allowed ranges.

Computer-assisted qualitative data analysis

The use of special computer software to assist qualitative analyses through creating, applying, and refining categories; tracing linkages between concepts; and making comparisons between cases and events.

Computer-assisted telephone interview (CATI)

A telephone interview in which a questionnaire is programmed into a computer, along with relevant skip patterns, and only valid entries are allowed; incorporates the tasks of interviewing, data entry, and some data cleaning.

Concept

A mental image that summarizes a set of similar observations, feelings, or ideas.

Conceptualization

The process of specifying what we mean by a term. In deductive research, conceptualization helps translate portions of an abstract theory into specific variables that can be used in testable hypotheses. In inductive research, conceptualization is an important part of the process used to make sense of related observations.

Concurrent validity

The type of validity that exists when scores on a measure are closely related to scores on a criterion measured at the same time.

Confidentiality

A provision of research, in which identifying information that could be used to link respondents to their responses is available only to designated research personnel for specific research needs.

Conflict of interest

When a researcher has a significant financial stake in the design or outcome of his or her own research.

Conflict theory

A social theory that identifies conflict between social groups as the primary force in society; understanding the bases and consequences of the conflict is key to understanding social processes.

Conjunctural research

Research that considers the complex combinations in which causal influences operate.

Constant

A number that has a fixed value in a given situation; a characteristic or value that does not change.

Construct validity

The type of validity that is established by showing that a measure is related to other measures as specified in a theory.

Constructivism

A methodology based on questioning belief in an external reality; emphasizes the importance of exploring the way in which different stakeholders in a social setting construct their beliefs.

Contamination

A source of causal invalidity that occurs when either the experimental or the comparison group is aware of the other group and is influenced in the posttest as a result.

Content analysis

A research method for systematically analyzing and making inferences from recorded human communication, including books, articles, poems, constitutions, speeches, and songs.

Content validity

The type of validity that exists when the full range of a concept’s meaning is covered by the measure.

Context

A set of interrelated circumstances that alters a relationship between other variables or social processes.

Context effects

When one or more survey questions influence how subsequent questions are interpreted.

Contextual effect

A variation in relationships of dependent with independent variables between geographic units or other social settings.

Contingent question

A question that is asked of only a subset of survey respondents.

Continuous measure

A measure with numbers indicating the values of variables as points on a continuum.

Control group

A comparison group that receives no treatment.

Convergent validity

An approach to construct validation; the type of validity achieved when one measure of a concept is associated with different types of measures of the same concept.

Conversation analysis

A qualitative method for analyzing the sequential organization and details of ordinary conversation.

Correlation coefficient

A summary statistic that varies from 0 to 1 or −1, with 0 indicating the absence of a linear relationship between two quantitative variables and 1 or −1 indicating that the relationship is completely described by the line representing the regression of the dependent variable on the independent variable.

Correlational analysis

A statistical technique that summarizes the strength of a relationship between two quantitative variables in terms of its adherence to a linear pattern.

Cost–benefit analysis

A type of evaluation research that compares program costs with the economic value of program benefits.

Cost–effectiveness analysis

A type of evaluation research that compares program costs with actual program outcomes.

Counterfactual

The situation that would have occurred if the subjects who were exposed to the treatment were not exposed but otherwise had had identical experiences to those they underwent during the experiment.

Cover letter

The letter sent with a mailed questionnaire that explains the survey’s purpose and auspices and encourages the respondent to participate.

Covert observer

A role in participant observation in which the researcher does not participate in group activities and is not publicly defined as a researcher.

Criterion validity

The type of validity that is established by comparing the scores obtained on the measure being validated with those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

Cronbach's alpha

A statistic commonly used to measure interitem reliability.

Cross-population generalizability (external validity)

Exists when findings about one group, population, or setting hold true for other groups, populations, or settings.

Cross-sectional comparative research

Research comparing data from one time period between two or more nations.

Cross-sectional research design

A study in which data are collected at only one point in time.

Cross-tabulation (crosstab) or contingency table

In the simplest case, a bivariate (two-variable) distribution, showing the distribution of one variable for each category of another variable; can be elaborated using three or more variables.

Curvilinear

Any pattern of association between two quantitative variables that does not involve a regular increase or decrease.

Data cleaning

The process of checking data for errors after the data have been entered in a computer file.

Data entry

The process of typing (word processing) or otherwise transferring data on survey or other instruments into a computer file.

Data visualization

A rapidly developing field that focuses on how to represent and present data to facilitate understanding.

Debriefing

A researcher's informing subjects after an experiment about the experiment's purposes and methods and evaluating subjects' personal reactions to the experiment.

Deception

Used in social experiments to create more "realistic" treatments in which the true purpose of the research is not disclosed to participants, often within the confines of a laboratory.

Deductive research

The type of research in which a specific expectation is deduced from a general premise and is then tested.

Demography
The statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time.

Demoralization
A type of contamination in experimental and quasi-experimental designs that occurs when control group members feel they have been left out of some valuable treatment and perform worse as a result.

Dependent variable
A variable that is hypothesized to vary depending on, or under the influence of, another variable.

Descriptive research
Research in which social phenomena are defined and described.

Descriptive statistics
Statistics used to describe the distribution of and relationship between variables.

Dichotomy
Variable having only two values.

Differential attrition (mortality)
A problem that occurs in experiments when comparison groups become different because subjects are more likely to drop out of one of the groups for various reasons.

Digital ethnography
The use of ethnographic methods to study online communities; also termed *netnography*, *cyberethnography,* and *virtual ethnography*.

Direction of association
A pattern in a relationship between two variables—the values of variables tend to change consistently in relation to change on the other variable; the direction of association can be either positive or negative.

Discrete measure
A measure that classifies cases in distinct categories.

Discriminant validity
An approach to construct validation; the scores on the measure to be validated are compared with scores on another measure of the same variable and to scores on

variables that measure different but related concepts. Discriminant validity is achieved if the measure to be validated is related most strongly to its comparison measure and less so to the measures of other concepts.

Disproportionate stratified sampling

A sampling method in which elements are selected from strata in different proportions from those that appear in the population.

Double-barreled question

A single survey question that actually asks two questions but allows only one answer.

Double-blind procedure

An experimental method in which neither the subjects nor the staff delivering experimental treatments know which subjects are getting the treatment and which are receiving a placebo.

Double negative

A question or statement that contains two negatives, which can muddy the meaning of the question.

Ecological fallacy

An error in reasoning in which incorrect conclusions about individual-level processes are drawn from group-level data.

Effect size

A standardized measure of association—often the difference between the mean of the experimental group and the mean of the control group on the dependent variable, adjusted for the average variability in the two groups.

Efficiency analysis

A type of evaluation research that compares program costs with program effects. It can be either a cost–benefit analysis or a cost–effectiveness analysis.

Elaboration analysis

The process of introducing a third variable into an analysis to better understand—to elaborate—the bivariate (two-variable) relationship under consideration. Additional control variables also can be introduced.

Electronic survey

A survey that is sent and answered by computer, either through e-mail or on the web.

Elements

The individual members of the population whose characteristics are to be measured.

Embedded mixed-methods design

Qualitative and quantitative methods are used concurrently in the research but one is given priority.

Emergence

The appearance of phenomena at a group level that cannot be explained by the properties of individuals within the group; emergence implies phenomena that are more than “the sum of their parts.”

Emic focus

Representing a setting with the participants’ terms and from their viewpoint.

Empirical generalization

A statement that describes patterns found in data.

Enumeration units

Units that contain one or more elements and that are listed in a sampling frame.

Ethnography

The study of a culture or cultures that some group of people shares, using participant observation over an extended period.

Ethnomethodology

A qualitative research method focused on the way that participants in a social setting create and sustain a sense of reality.

Etic focus

Representing a setting with the researchers’ terms and from their viewpoint.

Evaluability assessment

A type of evaluation research conducted to determine whether it is feasible to evaluate a program’s effects within the available time and resources.

Evaluation research

Research that describes or identifies the impact of social policies and programs.

Event-based design (cohort study)

A type of longitudinal study in which data are collected at two or more points in time from individuals in a cohort.

Event-structure analysis

A systematic method of developing a causal diagram showing the structure of action underlying some chronology of events; the result is an idiographic causal explanation.

Ex post facto control group design

A nonexperimental design in which comparison groups are selected after the treatment, program, or other variation in the independent variable has occurred, but when the participants were able to choose the group in which they participated. Often confused with a quasi-experimental design.

Exhaustive

A question's response choices are exhaustive when they cover all possible responses.

Expectancies of experimental staff

A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects results from the positive expectancies of the staff who are delivering the treatment rather than from the treatment itself; also called a *self-fulfilling prophecy*.

Experience sampling method (ESM)

A technique for drawing a representative sample of everyday activities, thoughts, and experiences. Participants carry a pager and are beeped at random times over several days or weeks; on hearing the beep, participants complete a report designed by the researcher.

Experimental group

In an experiment, the group of subjects that receives the treatment or experimental manipulation.

Explanatory research

Research that seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon.

Exploratory research

Research that seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them.

Extraneous variable

A variable that influences both the independent and dependent variables, creating a spurious association between them that disappears when the extraneous variable is controlled.

Face validity

The type of validity that exists when an inspection of items used to measure a concept suggests that they are appropriate "on their face."

Factorial survey

A survey in which randomly selected subsets of respondents are asked different questions, or are asked to respond to different vignettes, to determine the causal effect of the variables represented by these differences.

Federal Policy for the Protection of Human Subjects (also known as the Common Rule)

Specific regulations adopted in 1991 by the U.S. Department of Health and Human Services and the U.S. Food and Drug Administration that were based on the principles of the *Belmont Report*. Revised in 2017, with new requirements in effect, January 2018.

Feedback

Information about service delivery system outputs, outcomes, or operations that can guide program input.

Fence-sitters

Survey respondents who see themselves as being neutral on an issue and choose a middle (neutral) response that is offered.

Field experiment

A study using an experimental design that is conducted in a real-world setting.

Field notes

Notes that describe what has been observed, heard, or otherwise experienced in a participant observation study. These notes usually are written after the observational session.

Field research

Research in which natural social processes are studied as they happen and left relatively undisturbed.

Field researcher

A researcher who uses qualitative methods to conduct research in the field.

Filter question

A survey question used to identify a subset of respondents who then are asked other questions.

Fixed-sample panel design (panel study)

A type of longitudinal study in which data are collected from the same individuals—the panel—at two or more points in time. In another type of panel design, panel members who leave are replaced with new members.

Floaters

Survey respondents who provide an opinion on a topic in response to a closed-ended question that does not include a “Don’t know” option, but who will choose “Don’t know” if it is available.

Focus groups
A qualitative method that involves unstructured group interviews in which the focus group leader actively encourages discussion among participants on the topics of interest.

Forced-choice questions
Closed-ended survey questions that do not include “Don’t know” as an explicit response choice.

Formative evaluation
Process evaluation that is used to shape and refine program operations.

Frequency distribution
A numerical display showing the number of cases, and usually the percentage of cases (the relative frequencies), corresponding to each value or group of values of a variable.

Frequency polygon
A graphic for quantitative variables in which a continuous line connects data points representing the variable’s distribution.

Front matter
The section of an applied research report that includes an executive summary, an abstract, and a table of contents.

Functionalism
A social theory that explains social patterns in terms of their consequences for society as a whole and that emphasizes the interdependence of social institutions and their common interest in maintaining the social order.

Gamma
A measure of association that is sometimes used in cross-tabular analysis.

Gatekeeper
A person in a field setting who can grant researchers access to the setting.

Generalizability
Exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify.

Grand tour question
A broad question at the start of an interview that seeks to engage the respondent in

the topic of interest.

Grounded theory

A systematic theory developed inductively, based on observations that are summarized into conceptual categories, reevaluated in the research setting, and gradually refined and linked to other conceptual categories.

Group-administered survey

A survey that is completed by individual respondents who are assembled in a group.

Hawthorne effect

A type of contamination in research designs that occurs when members of the treatment group change relative to the dependent variable because their participation in the study makes them feel special.

Health Insurance Portability and Accountability Act (HIPAA)

Congressional legislation passed in 1996 that creates stringent regulations for the protection of health care data.

Hermeneutic circle

A representation of the dialectical process in which the researcher obtains information from multiple stakeholders in a setting, refines his or her understanding of the setting, and then tests that understanding with successive respondents.

Histogram

A graphic for quantitative variables in which the variable's distribution is displayed with adjacent bars.

Historical events research

Research in which social events are studied at one past time period.

Historical process research

Research in which historical processes are studied over a long time.

History effect

A source of causal invalidity that occurs when events external to the study influence posttest scores; also called an effect of *external events*.

Holistic research

Research concerned with the context in which events occurred and the interrelations between different events and processes.

Hypothesis

A tentative statement about empirical reality, involving a relationship between two or more variables.

Idiographic causal explanation

An explanation that identifies the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event; may be termed an *individualist* or *historicist* explanation.

Idiosyncratic (or "random") errors

Errors that affect individuals or other cases in unique ways that are unlikely to be repeated in just the same way.

Idiosyncratic variation

Variation in responses to questions that is caused by individuals' reactions to particular words or ideas in the question instead of by variation in the concept that the question is intended to measure.

Illogical reasoning

When we prematurely jump to conclusions or argue on the basis of invalid assumptions.

Impact evaluation (or analysis)

Analysis of the extent to which a treatment or other service has an effect; also known as summative evaluation.

Inaccurate observation

An observation based on faulty perceptions of empirical reality.

Independent variable

A variable that is hypothesized to cause, or lead to, variation in another variable.

Index

The sum or average of responses to a set of questions about a concept.

Indicator

The question or other operation used to indicate the value of cases on a variable.

Inductive research

The type of research in which general conclusions are drawn from specific data.

Inferential statistics

A mathematical tool for estimating how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

In-person interview

A survey in which an interviewer questions respondents face-to-face and records their answers.

Inputs

The resources, raw materials, clients, and staff that go into a program.

Institutional review board (IRB)

A group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to human subjects.

Integrated mixed-methods design

Qualitative and quantitative methods are used concurrently and both are given equal importance.

Integrative approach

An orientation to evaluation research that expects researchers to respond to the concerns of people involved with the program—stakeholders—as well as to the standards and goals of the social scientific community.

Intensive (in-depth) interviewing

A qualitative method that involves open-ended, relatively unstructured questioning in which the interviewer seeks in-depth information on the interviewee's feeling, experiences, and perceptions (Lofland and Lofland 1984:12).

Intent-to-treat analysis

When analysis of the effect of a treatment on outcomes in an experimental design compares outcomes for all those who were assigned to the treatment group with outcomes for all those who were assigned to the control group, whether or not participants remained in the treatment group.

Interactive voice response (IVR)

A survey in which respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software.

Intercoder reliability

When the same codes are entered by different coders who are recording the same data.

Interitem reliability

An approach that calculates reliability based on the correlation among multiple items used to measure a single concept; also known as internal consistency.

Interobserver reliability

When similar measurements are obtained by different observers rating the same persons, events, or places.

Interpretive questions

Questions included in a questionnaire or interview schedule to help explain answers to other important questions.

Interpretivism

The belief that the subjective meanings people give to their experiences are a key focus for social science research without assuming that social processes can be identified objectively.

Interquartile range

The range in a distribution between the end of the first quartile and the beginning of the third quartile.

Intersubjective agreement

Agreement between scientists about the nature of reality; often upheld as a more reasonable goal for science than certainty about an objective reality.

Inter-university Consortium for Political and Social Research (ICPSR)

The academic consortium that archives data sets online from major surveys and other social science research and makes them available for analysis by others.

Interval level of measurement

A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.

Interval–ratio level of measurement

A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but may not have an absolute, or fixed, zero point.

Intervening variable

See mediator.

Interview schedule

The survey instrument containing the questions asked by the interviewer in an in-person or phone survey.

Intrarater (or intraobserver) reliability

Consistency of ratings by an observer of an unchanging phenomenon at two or more points in time.

Jottings

Brief notes written in the field about highlights of an observation period.

Justice

The ethical principle of distributing benefits and risks of research fairly in research involving human subjects that was included in the *Belmont Report*.

Key informant

An insider who is willing and able to provide a field researcher with superior access and information, including answers to questions that arise in the course of the research.

Labeled unipolar response options

Response choices for a survey question that use words to identify categories ranging from low to high (or high to low).

Level of measurement

The mathematical precision with which the values of a variable can be expressed. The nominal level of measurement, which is qualitative, has no mathematical interpretation; the quantitative levels of measurement—ordinal, interval, and ratio—are progressively more precise mathematically.

Likert item

A statement followed by response choices ranging from "strongly agree" to "strongly disagree."

Longitudinal research design

A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.

Mailed survey

A survey involving a mailed questionnaire to be completed by the respondent.

Marginal distribution

The summary distributions in the margins of a cross-tabulation that correspond to the frequency distribution of the row variable and of the column variable.

Matching

A procedure for equating the characteristics of individuals in different comparison groups in an experiment. Matching can be done on either an individual or an aggregate basis. For individual matching, individuals who are similar in key characteristics are paired before assignment, and then the two members of each pair are assigned to the two groups. For aggregate matching, also termed *blocking*, groups that are chosen for comparison are similar in the distribution of key characteristics.

Matrix

A form on which particular features of multiple cases or instances can be recorded systematically so that a qualitative data analyst can examine them later.

Matrix questions

A series of survey questions that concern a common theme and that have the same response choices.

Mean

The arithmetic, or weighted, average, computed by adding the value of all the cases and dividing by the total number of cases.

Measure of association

A type of descriptive statistic that summarizes the strength of an association.

Measurement validity

Exists when a measure measures what we think it measures.

Measurement

The process of linking abstract concepts to empirical indicants.

Mechanism

A discernible process that creates a causal connection between two variables.

Median

The position average, or the point that divides a distribution in half (the 50th percentile).

Mediator

A variable involved in a causal mechanism (intervening variable).

Meta-analysis

The quantitative analysis of findings from multiple studies.

Method of agreement

A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that agree on an outcome variable also agree on the value of the variable hypothesized to have a causal effect, although they differ on other variables.

Method of difference

A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that differ on an outcome variable also differ on the value of the variable hypothesized to have a causal effect, although they agree on other variables.

Milgram's obedience experiments

Experiments begun in 1960 at Yale University by psychologist Stanley Milgram to determine the likelihood of people following orders from an authority despite their own sentiments; widely cited as helping to understand the emergence of phenomena such as Nazism and mass cults.

Mixed methods

Research that combines qualitative and quantitative methods in an investigation of the same or related research question(s).

Mixed-mode survey

A survey that is conducted by more than one method, allowing the strengths of one survey design to compensate for the weaknesses of another and maximizing the likelihood of securing data from different types of respondents; for example, nonrespondents in a mailed survey may be interviewed in person or over the phone.

Mode

The most frequent value in a distribution; also termed the *probability average*.

Moderator

A variable that identifies a context for the effect of other variables.

Monotonic

A pattern of association in which the value of cases on one variable increases or decreases fairly regularly across the categories of another variable.

Multiple group before-and-after design

A type of quasi-experimental design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.

Multitrait–multimethod matrix

A method of evaluating the validity of measures by determining whether measuring the same phenomenon with different methods leads to convergent results and measuring different phenomena with the same method leads to divergent results.

Mutually exclusive

A question's response choices are mutually exclusive when every case can be classified as having only one attribute (or value).

Narrative analysis

A form of qualitative analysis in which the analyst focuses on how respondents impose order on the flow of experience in their lives and thus make sense of events and actions in which they have participated.

Narrative explanation

An idiographic causal explanation that involves developing a narrative of events and processes that indicate a chain of causes and effects.

Needs assessment

A type of evaluation research that attempts to determine the needs of some population that might be met with a social program.

Ngrams

Frequency graphs produced by Google's database of all words printed in more than one third of the world's books over time (with coverage still expanding).

Nominal level of measurement

Variables whose values have no mathematical interpretation; they vary in kind or quality, but not in amount.

Nomothetic causal explanation

An explanation that identifies common influences on a number of cases or events.

Nonequivalent control group design

A quasi-experimental design in which experimental and comparison groups are designated before the treatment occurs but are not created by random assignment.

Nonprobability sampling method

A sampling method in which the probability of selection of population elements is unknown.

Nonrespondents

People or other entities who do not participate in a study although they are selected for the sample.

Nonspuriousness

A criterion for establishing a causal relation between two variables; when a relationship between two variables is not caused by variation in a third variable.

Normal distribution

A symmetric, bell-shaped distribution that results from chance variation around a central value.

Normal science

The gradual, incremental research conducted by scientists within the prevailing scientific paradigm.

Nuremberg War Crime Trials

The International Military Tribunal held by the victorious Allies after World War II in Nuremberg, Germany, that exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science."

Office for Protection From Research Risks, National Institutes of Health

The office in the U.S. Department of Health and Human Services that provides leadership and supervision about the protection of the rights, welfare, and well-being of subjects involved in research conducted or supported by DHHS, including monitoring IRBs.

Omnibus survey

A survey that covers a range of topics of interest to different social scientists.

Open-ended question

A survey question to which the respondent replies in his or her own words, either by writing or by talking.

Operationalization

The process of specifying the measures that will indicate the value of cases on a variable.

Oral history

Data collected through intensive interviews with participants in past events.

Ordinal level of measurement

A measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.

Outcomes

The impact of the program process on the cases processed.

Outlier

An exceptionally high or low value in a distribution.

Outputs

The services delivered or new products produced by the program process.

Overgeneralization

When we unjustifiably conclude that what is true for some cases is true for all cases.

Panel study

See fixed-sample panel design.

Paradigm wars

The intense debate from the 1970s to the 1990s between social scientists over the value of positivist and interpretivist/constructivist research philosophies; also see *scientific paradigm*.

Participant observation

A qualitative method for gathering data that involves developing a sustained relationship with people while they go about their normal activities.

Participant observer

A researcher who gathers data through participating and observing in a setting where he or she develops a sustained relationship with people while they go about their normal activities. The term *participant observer* is often used to refer to a continuum of possible roles, from complete observation, in which the researcher does not participate along with others in group activities, to complete participation, in which the researcher participates without publicly acknowledging being an observer.

Participatory action research (PAR)

A type of research in which the researcher involves members of the population to be studied as active participants throughout the research process, from the selection of a research focus to the reporting of research results and efforts to make changes based on the research; also termed *community-based participatory research*.

Part–whole question effects

When responses to a general or summary survey question about a topic are influenced by responses to an earlier, more specific question about that topic.

Percentages

Relative frequencies, computed by dividing the frequency of cases in a particular category by the total number of cases and then multiplying by 100.

Periodicity

A sequence of elements (in a list to be sampled) that varies in some regular, periodic pattern.

Phone survey

A survey in which interviewers question respondents over the phone and then record their answers.

Photo voice

A method in which research participants take pictures of their everyday surroundings with cameras the researcher distributes, and then meet in a group with the researcher to discuss the pictures' meaning.

Placebo effect

A source of treatment misidentification that can occur when subjects receive a fake "treatment" they think is beneficial and improve because of that expectation even though they did not receive the actual treatment or received a treatment that had no real effect.

Plagiarism

Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.

Policy research

A process in which research results are used to provide policy actors with recommendations for action that are based on empirical evidence and careful reasoning.

Population

The entire set of individuals or other entities to which study findings are to be generalized.

Population parameter

The value of a statistic, such as a mean, computed using the data for the entire population; a sample statistic is an estimate of a population parameter.

Positivism

The belief, shared by most scientists, that there is a reality that exists quite apart from our own perception of it, that it can be understood through observation, and that it follows general laws.

Postpositivism

A philosophical view that modifies the positivist premise of an external, objective reality by recognizing its complexity, the limitations of human observers, and therefore the impossibility of developing more than a partial understanding of reality.

Posttest

In experimental research, the measurement of an outcome (dependent) variable after an experimental intervention or after a presumed independent variable has changed for some other reason.

Precoding

The process through which a questionnaire or other survey form is prepared so that a number represents every response choice, and respondents are instructed to indicate their response to a question by checking a number.

Predictive validity

The type of validity that exists when a measure predicts scores on a criterion measured in the future.

Pretest

In experimental research, the measurement of an outcome (dependent) variable before an experimental intervention or change in a presumed independent variable for some other reason. The pretest is exactly the same "test" as the posttest, but it is administered at a different time.

Probability average

See mode.

Probability of selection

The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all elements of a population, the probability that any particular element will be selected is 1.0. If half the elements in the population are sampled on the basis of chance (say, by tossing a coin), the probability of selection for each element is one half, or .5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

Probability sampling method

A sampling method that relies on a random, or chance, selection method so that the probability of selection of population elements is known.

Procedural justice theory

A social theory that predicts that people will obey the law from a sense of obligation that flows from seeing legal authorities as moral and legitimate.

Process consent

An interpretation of the ethical standard of voluntary consent that allows participants to change their decision about participating at any point by requiring that the researcher check with participants at each stage of the project about their willingness to continue in the project.

Process evaluation

Evaluation research that investigates the process of service delivery.

Program process

The complete treatment or service delivered by the program.

Program theory

A descriptive or prescriptive model of how a program operates and produces effects.

Progressive focusing

The process by which a qualitative analyst interacts with the data and gradually refines his or her focus.

Proportionate stratified sampling

A sampling method in which elements are selected from strata in exact proportion to their representation in the population.

Pseudoscience

Claims presented so that they appear scientific even though they lack supporting evidence and plausibility.

Publication bias

The greater likelihood of publication of articles reporting tests of hypotheses that were supported.

Purposive sampling

A nonprobability sampling method in which elements are selected for a purpose, usually because of their unique position.

Qualitative comparative analysis (QCA)

A systematic type of qualitative analysis that identifies the combination of factors that had to be present across multiple cases to produce a particular outcome.

Qualitative methods

Methods such as participant observation, intensive interviewing, and focus groups that are designed to capture social life as participants experience it rather than in categories predetermined by the researcher. These methods rely on written or spoken words or observations that do not often have a direct numerical interpretation and typically involve exploratory research questions, inductive reasoning, an orientation to social context and human subjectivity, and the meanings attached by participants to events and to their lives.

Quantitative methods

Methods such as surveys and experiments that record variation in social life in terms of categories that vary in amount. Data that are treated as quantitative are either numbers or attributes that can be ordered by magnitude.

Quartiles

The points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases.

Quasi-experimental design

A research design in which there is a comparison group that is comparable to the experimental group in critical ways, but subjects are not randomly assigned to the comparison and experimental groups.

Questionnaire

The survey instrument containing the questions in a self-administered survey.

Quota sampling

A nonprobability sampling method in which elements are selected to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Random assignment

A procedure by which each experimental subject is placed in a group randomly.

Random digit dialing

The random machine-dialing of numbers within designated phone prefixes, which creates a random sample for phone surveys.

Random number table

A table containing lists of numbers that are ordered solely on the basis of chance; it is used for drawing a random sample.

Random sampling

A method of sampling that relies on a random, or chance, selection method so that every element of the sampling frame has a known probability of being selected.

Random sampling error (chance sampling error)

Differences between the population and the sample that are due only to chance factors (random error), not to systematic sampling error. Random sampling error may or may not result in an unrepresentative sample. The magnitude of sampling error resulting from chance factors can be estimated statistically.

Random variation

See idiosyncratic variation.

Randomization

The random assignment of cases, as by the toss of a coin.

Randomized comparative change design

The classic true experimental design in which subjects are assigned randomly to two groups; both these groups receive a pretest, then one group receives the experimental intervention, and then both groups receive a posttest. Also known as a *pretest–posttest control group design*.

Randomized comparative posttest design

A true experimental design in which subjects are assigned randomly to two groups—one group then receives the experimental intervention and both groups receive a posttest; there is no pretest. Also known as a *posttest-only control group design*.

Range

The true upper limit in a distribution minus the true lower limit (or the highest rounded value minus the lowest rounded value, plus one).

Ratio level of measurement

A measurement of a variable in which the numbers indicating a variable's values represent fixed measuring units and an absolute zero point.

Rational choice theory

A social theory that explains individual action with the principle that actors choose actions that maximize their gains from taking that action.

Reactive effects

The changes in individual or group behavior that result from being observed or otherwise studied.

Reductionist fallacy (reductionism)

An error in reasoning that occurs when incorrect conclusions about group-level processes are based on individual-level data; also known as an individualist fallacy.

Reflexivity

Sensitivity of and adaptation by the researcher to his or her influence in the research setting.

Regression analysis

A statistical technique for characterizing the pattern of a relationship between two quantitative variables in terms of a linear equation and for summarizing the strength of this relationship in terms of its deviation from that linear pattern.

Regression effect

A source of causal invalidity that occurs when subjects who are chosen for a study because of their extreme scores on the dependent variable become less extreme on the posttest because of natural cyclical or episodic change in the variable.

Regression–discontinuity design

A quasi-experimental design in which individuals are assigned to a treatment and a comparison group solely on the basis of a cutoff score on some assignment variable, and then treatment effects are identified by a discontinuity in the regression line that displays the relation between the outcome and the assignment variable at the cutoff score.

Reliability

When a measurement procedure yields consistent scores when the phenomenon being measured is not changing.

Reliability measures

Statistics that summarize the consistency among a set of measures; Cronbach's alpha is the most common measure of the reliability of a set of items included in an index.

Repeated cross-sectional design (trend study)

A type of longitudinal study in which data are collected at two or more points in time from different samples of the same population.

Repeated measures panel design

A quasi-experimental design consisting of several pretest and posttest observations of the same group.

Replacement sampling

A method of sampling in which sample elements are returned to the sampling frame after being selected, so they may be sampled again. Random samples may be selected with or without replacement.

Replications

Repetitions of a study using the same research methods to answer the same research question.

Representative sample

A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Research circle

A diagram of the elements of the research process, including theories, hypotheses, data collection, and data analysis.

Research program with mixed methods

Qualitative and quantitative methods are used in sequence and are given equal priority.

Resistance to change

The reluctance to change our ideas in light of new information.

Respect for persons

The ethical principle of treating persons as autonomous agents and protecting those with diminished autonomy in research involving human subjects that was included in the *Belmont Report*.

Responsive evaluation

See stakeholder approach.

Reverse outlining

Outlining the sections in an already written draft of a paper or report to improve its organization in the next draft.

Sample

A subset of a population that is used to study the population as a whole.

Sample generalizability

Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

Sample statistic

The value of a statistic, such as a mean, computed from sample data.

Sampling error

Any difference between the characteristics of a sample and the characteristics of a population; the larger the sampling error, the less representative the sample.

Sampling frame

A list of all elements or other units containing the elements in a population.

Sampling interval

The number of cases from one sampled case to another in a systematic random sample.

Sampling units

Units listed at each stage of a multistage sampling design.

Saturation point

The point at which subject selection is ended in intensive interviewing, when new interviews seem to yield little additional information.

Science

A set of logical, systematic, documented methods for investigating nature and natural processes; the knowledge produced by these investigations.

Scientific paradigm

A set of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings.

Secondary data

Previously collected data that are used in a new analysis.

Secondary data analysis

The method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data.

Selection bias

A source of internal (causal) invalidity that occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.

Selective distribution of benefits

An ethical issue about how much researchers can influence the benefits subjects receive as part of the treatment being studied in a field experiment.

Selective observation

Choosing to look only at things that are in line with our preferences or beliefs.

Serendipitous findings

Unexpected patterns in data, which stimulate new explanations, insights, or theoretical approaches.

Simple random sampling

A method of sampling in which every sample element is selected only on the basis of chance, through a random process.

Skewness

The extent to which cases are clustered more at one or the other end of the distribution of a quantitative variable rather than in a symmetric pattern around its center. Skew can be positive (a right skew), with the number of cases tapering off in the positive direction, or negative (a left skew), with the number of cases tapering off in the negative direction.

Skip pattern

The unique combination of questions created in a survey by filter questions and contingent questions.

Snowball sampling

A nonprobability sampling method in which sample elements are selected as they are identified by successive informants or interviewees.

Social desirability bias

The tendency to "agree" with a statement just to avoid seeming disagreeable.

Social research question

A question about the social world that is answered through the collection and analysis of firsthand, verifiable, empirical data.

Social science

The use of scientific methods to investigate individuals, societies, and social processes; the knowledge produced by these investigations.

Social science (researcher) approach

An orientation to evaluation research that expects researchers to emphasize the importance of researcher expertise and maintenance of autonomy from program stakeholders.

Solomon four-group design

A type of experimental design that combines a randomized pretest–posttest control group design with a randomized posttest-only design, resulting in two experimental groups and two comparison groups.

Specification

A type of relationship involving three or more variables in which the association between the independent and dependent variables varies across the categories of one or more other control variables.

Split-ballot design

Unique questions or other modifications in a survey administered to randomly selected subsets of the total survey sample, so that more questions can be included in the entire survey or so that responses to different question versions can be compared.

Split-half reliability

Reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.

Spurious relationship

A relationship between two variables that is caused by variation in a third variable.

Staged mixed-methods design

Qualitative and quantitative methods are used in sequence in the research and one is given priority.

Stakeholder approach

An orientation to evaluation research that expects researchers to be responsive primarily to the people involved with the program; also termed responsive evaluation.

Stakeholders

Individuals and groups who have some basis of concern with the program.

Standard deviation

The square root of the average squared deviation of each case from the mean.

Statistical control

A method in which one variable is held constant so that the relationship between two (or more) other variables can be assessed without the influence of variation in the control variable.

Statistical significance

The mathematical likelihood that an association is due to chance, judged by a criterion set by the analyst (often that the probability is less than 5 out of 100 or $p < .05$).

Stratified random sampling

A sampling method in which sample elements are selected separately from population strata that are identified in advance by the researcher.

Street-level bureaucrats

Officials who serve clients and have a high degree of discretion.

Subject fatigue

Problems caused by panel members growing weary of repeated interviews and dropping out of a study or becoming so used to answering the standard questions in the survey that they start giving stock or thoughtless answers.

Subtables

Tables describing the relationship between two variables within the discrete categories of one or more other control variables.

Summative evaluation

See impact evaluation (or analysis).

Survey pretest

A method of evaluating survey questions and procedures by testing them on a small sample of individuals like those to be included in the actual survey and then reviewing responses to the questions and reactions to the survey procedures.

Survey research

Research in which information is obtained from a sample of individuals through their responses to questions about themselves or others.

Symbolic interaction theory

A social theory that focuses on the symbolic nature of social interaction—how social interaction conveys meaning and promotes socialization.

Systematic bias

Overrepresentation or underrepresentation of some population characteristics in a sample resulting from the method used to select the sample; a sample shaped by systematic sampling error is a biased sample.

Systematic errors

Errors that result from factors that are not what an instrument is intended to measure and that affect individuals or other cases in ways that are likely to recur in just the same way, thus creating measurement bias.

Systematic observation

A strategy that increases the reliability of observational data by using explicit rules that standardize coding practices across observers.

Systematic random sampling

A method of sampling in which sample elements are selected from a list or from sequential files, with every *n*th element being selected after the first element is selected randomly within the first interval.

Systematic review

A literature review that "uses a specific and reproducible method to identify, select and appraise studies of a previously agreed level of quality (either including all studies or only those that pass a minimum quality threshold) that are relevant to a particular question" (Booth et al. 2016:11).

Tacit knowledge

In field research, a credible sense of understanding of social processes that reflects the researcher's awareness of participants' actions as well as their words, and of what they fail to state, feel deeply, and take for granted.

Target population

A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.

Tearoom Trade

The study by sociologist Laud Humphreys of men who engaged in homosexual behavior in public facilities, including subsequent later interviews in their homes after recording their license plate numbers; widely cited in discussions of the need for informed consent to research.

Temporal research

Research that accounts for the related series of events that unfold over time.

Test–retest reliability

A measurement showing that measures of a phenomenon at two points in time are highly correlated, if the phenomenon has not changed or has changed only as much as the measures have changed.

Theoretical sampling

A sampling method recommended for field researchers by Glaser and Strauss (1967). A theoretical sample is drawn in a sequential fashion, with settings or individuals selected for study as earlier observations or interviews indicate that these settings or individuals are influential.

Theory

A logically interrelated set of propositions about empirical reality.

Theory-driven evaluation

A program evaluation that is guided by a theory that specifies the process by which the program has an effect.

Thick description

A rich description that conveys a sense of what it is like from the standpoint of the natural actors in that setting.

Time order

A criterion for establishing a causal relation between two variables; the variation in the presumed cause (the independent variable) must occur before the variation in the presumed effect (the dependent variable).

Time series design

A quasi-experimental design consisting of many pretest and posttest observations of the same group over an extended period.

Trend study

See repeated cross-sectional design.

Triangulation

The use of multiple methods to study one research question; the use of two or more different measures of the same variable.

True experiment

An experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation; outcomes are measured in a posttest.

Tuskegee Study of Untreated Syphilis in the Negro Male

The U.S. Public Health Service study of the "natural" course of syphilis that followed 399 low-income African American men from the 1930s to 1972, without providing them with penicillin after the drug was discovered to treat the illness. The study was stopped after it was exposed in 1972, resulting in an out-of-court settlement and then, in 1997, an official public apology by President Bill Clinton.

Unbalanced response choices

When a fixed-choice survey question has a different number of positive and negative response choices.

Unimodal

A distribution of a variable in which there is only one value that is the most frequent.

Units of analysis

The level of social life on which a research question is focused, such as individuals, groups, towns, or nations.

Units of observation

The cases about which measures actually are obtained in a sample.

Unlabeled unipolar response options

Response choices for a survey question that use numbers to identify categories ranging from low to high (or high to low).

Unobtrusive measure

A measurement based on physical traces or other data that are collected without the knowledge or participation of the individuals or groups that generated the data.

Validity

The state that exists when statements or conclusions about empirical reality are correct.

Variability

The extent to which cases are spread out through the distribution or clustered in just one location.

Variable

A characteristic or property that can vary (take on different values or attributes).

Variable-oriented research

Research that focuses attention on variables representing particular aspects of the cases studied and then examines the relations between these variables across sets of cases.

Variance

A statistic that measures the variability of a distribution as the average squared deviation of each case from the mean.

Video ethnography

The use of audiovisual methods and editing techniques to record, analyze, and present one or more viewable social processes, actions, or events in interpretable segments.

Visual sociology

Sociological research in which the social world is "observed" and interpreted through photographs, films, and other images.

Web survey

A survey that is accessed and responded to on the World Wide Web.

Zimbardo prison simulation study

The famous prison simulation study at Stanford University by psychologist Philip Zimbardo, designed to investigate the impact of social position on behavior—specifically, the impact of being either a guard or a prisoner in a "total institution"; widely cited as demonstrating the likelihood of emergence of sadistic behavior in guards.

Bibliography

Abbey, Antonia. 2002. “Alcohol-Related Sexual Assault: A Common Problem Among College Students.” Journal of Studies on Alcohol, Suppl. 14:118–128.

Abbott, Andrew. 1992. “From Causes to Events: Notes on Narrative Positivism.” Sociological Methods and Research 20(May): 428–455.

Abbott, Andrew. 1994. “History and Sociology: The Lost Synthesis.” Pp. 77–112 in Engaging the Past: The Uses of History Across the Social Sciences, edited by Eric H. Monkkonen. Durham, NC: Duke University Press.

Abel, David. 2000. “Census May Fall Short at Colleges.” The Boston Sunday Globe, March 26, pp. B1, B4.

Abernathy, David. 2017. Using Geodata and Geolocation in the Social Sciences: Mapping Our Connected World. Thousand Oaks, CA: SAGE.

Abma, Tineke A. 2005. “Responsive Evaluation: Its Meaning and Special Contribution to Health Promotion.” Evaluation and Program Planning 28:279–289.

Abrutyn, Seth and Anna S. Mueller. 2014. “Are Suicidal Behaviors Contagious in Adolescence? Using Longitudinal Data to Examine Suicide Suggestion.” American Sociological Review 79:211–227.

Adair, G., T. W. Dushenko, and R. C. L. Lindsay. 1985. “Ethical Regulations and Their Impact on Research Practice.” American Psychologist 40:59–72.

Aguirre, Adalberto Jr. and David V. Baker. 1993. “Racial Prejudice and the Death Penalty: A Research Note.” Social Justice 20:150–155.

Aiden, Erez and Jean-Baptiste Michel. 2013. Uncharted: Big Data as a Lens on Human Culture. New York: Riverhead Books.

Akyüz, Aygül, Tülay Yavan, Gönül Şahiner, and Ayşe Kılıç. 2012. “Domestic Violence and Woman’s Reproductive Health: A Review of the Literature.” Aggression and Violent Behavior 17:514–518.

Aldrich, James O. and Hilda M. Rodriguez. 2013. Building SPSS Graphs to Understand Data. Thousand Oaks, CA: SAGE.

Alfred, Randall. 1976. “The Church of Satan.” Pp. 180–202 in The New Religious Consciousness, edited by Charles Glock and Robert Bellah. Berkeley: University of California Press.

Alise, Mark A. and Charles Teddlie. 2010. “A Continuation of the Paradigm Wars? Prevalence Rates of Methodological Approaches Across the Social/Behavioral Sciences.” Journal of Mixed Methods Research 4:103–126.

Alkire, Sabina, José Manuel Roche, Maria Emma Santos, and Suman Seth. 2011. “Multidimensional Poverty Index 2011: Brief Methodological Note.” Oxford, England: Oxford Poverty & Human Development Initiative (OPHI), Oxford Department of International Development. Retrieved from www.ophi.org.uk/wp-content/uploads/MPI_2011_Methodology_Note_4-11-2011_1500.pdf?79d835

Allport, Gordon. 1954. The Nature of Prejudice. Cambridge, MA: Addison-Wesley.

Altheide, David L. and John M. Johnson. 1994. “Criteria for Assessing Interpretive Validity in Qualitative Research.” Pp. 485–499 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Altman, Lawrence K. 1998. “Getting It Right on the Facts of Death.” The New York Times, December 22, p. D7.

Alwin, Duane F. and Jon A. Krosnick. 1991. “The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes.” Sociological Methods & Research 20:139–181.

Ambwani, Suman and Jaine Strauss. 2007. “Love Thyself Before Loving Others? A Qualitative and Quantitative Analysis of Gender Differences in Body Image and Romantic Love.” Sex Roles 56:13–21.

American Association for Public Opinion Research (AAPOR). 2014. “Nonresponse in RDD Cell Phone Surveys.” Retrieved from www.aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force_Report&Template=/CM/ContentDisplay.cfm&Conte

American Medical Association (AMA). 2011. AMA Code of Medical Ethics. Retrieved from www.ama-assn.org/ama/physician-resources/medical-ethics/code-medical-ethics.page?

American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA: American Psychiatric Publishing.

American Sociological Association (ASA). 1999. Code of Ethics and Policies and Procedures of the ASA Committee on Professional Ethics. Washington, DC: American Sociological Association.

American Sociological Association (ASA). n.d. Topic: Conflicts of Interest. Retrieved from http://www.asanet.org/teaching-learning/faculty/teaching-ethics-throughout-curriculum/topic-conflicts-interest

Anderson, Elijah. 1990. Streetwise: Race, Class, and Change in an Urban Community. Chicago: University of Chicago Press.

Anderson, Elijah. 1999. Code of the Street: Decency, Violence, and the Moral Life of the Inner City. New York: Norton.

Anderson, Elijah. 2003. “Jelly’s Place: An Ethnographic Memoir.” Symbolic Interaction 26:217–237.

Anderson, Monica and Andrew W. Perrin. 2016. “13% of Americans Don’t Use the Internet. Who Are They?” Pew Research Center, September 7. Retrieved from http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/

Anderson, Rick. 2017. “Cabell’s New Predatory Journal Blacklist: A Review.” The Scholarly Kitchen, July 25. Retrieved from https://scholarlykitchen.sspnet.org/2017/07/25/cabells-new-predatory-journal-blacklist-review/

Anderton, Douglas L., Richard E. Barrett, and Donald J. Bogue. 1997. The Population of the United States. New York: Free Press.

Andresen E., T. K. Malmstrom, F. D. Wolinsky, M. Schootman, J. P. Miller, and D. K. Miller. 2008. “Rating Neighborhoods for Older Adult Health: Results From the African American Health Study.” BMC Public Health 8:35. doi:10.1186/1471-2458-8-35

Aneshensel, Carol S. 2002. Theory-Based Data Analysis for the Social Sciences. Thousand Oaks, CA: SAGE.

Anspach, Renee R. 1991. “Everyday Methods for Assessing Organizational Effectiveness.” Social Problems 38(February):1–19.

Antaki, Charles. 2008. “Discourse Analysis and Conversation Analysis.” Pp. 431–446 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Applebaum, Robert, Valerie Wellin, Cary Kart, J. Scott Brown, Heather Menne, Farida Ejaz, and Keren Brown Wilson. 2007. Evaluation of Ohio’s Assisted Living Medicaid Waiver Program: Final Summary Report. Oxford, OH: Scripps Gerontology Center and Miami University.

Archibold, Randal C. 2010. “On Border Violence, Truth Pales Compared to Ideas.” The New York Times, June 19. Retrieved from www.nytimes.com

Armas, Genaro C. 2002. “Government Won’t File Appeal in Census Case.” The Boston

Globe, November 23, p. A1.

Armstrong, Karen. 2008. “Ethnography and Audience.” Pp. 54–63 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Arthur, Michael W., John S. Briney, J. David Hawkins, Robert D. Abbott, Blair L. Brooke-Weiss, and Richard F. Catalano. 2007. “Measuring Risk and Protection in Communities Using the Communities That Care Youth Survey.” Evaluation and Program Planning 30:197–211.

Arwood, Tracy and Sangeeta Panicker. 2007. “Assessing Risk in Social and Behavioral Sciences.” Collaborative Institutional Training Initiative. Retrieved from https://www.citiprogram.org/members/learners

Asch, Solomon E. 1958. “Interpersonal Influence.” Pp. 174–183 in Readings in Social Psychology, 3rd ed., edited by Eleanor Maccoby, Theodore Newcomb, and Eugene Hartley. New York: Holt, Rinehart & Winston.

Aseltine, Robert H. Jr. and Ronald C. Kessler. 1993. “Marital Disruption and Depression in a Community Sample.” Journal of Health and Social Behavior 34(September):237–251.

Ayhan, H. Öztaş. 2001. “Gender Bias in Agricultural Surveys Statistics by Gender: Measures to Reduce Gender Bias in Agricultural Surveys.” International Statistical Review 69:447–460.

Babor, Thomas F., Robert S. Stephens, and G. Alan Marlatt. 1987. “Verbal Report Methods in Clinical Research on Alcoholism: Response Bias and Its Minimization.” Journal of Studies on Alcohol 48(5):410–424.

Bagadiya, Jimit. 2017. “151 Amazing Social Media Statistics You Should Know in 2017.” Social Pilot. Retrieved from https://www.socialpilot.co/blog/151-amazing-social-media-statistics-know-2017

Bail, Christopher A. 2008. "The Configuration of Symbolic Boundaries Against Immigrants in Europe." American Sociological Review 73:37–59.

Bainbridge, William Sims. 1989. Survey Research: A Computer-Assisted Introduction. Belmont, CA: Wadsworth.

Ball, Richard A. and G. David. Curry. 1995. "The Logic of Definition in Criminology: Purposes and Methods for Defining 'Gangs.'" Criminology 33:225–245.

Banks, J. A. 1972. "Historical Sociology and the Study of Population." Pp. 55–70 in Population and Social Change, edited by D. V. Glass and Roger Revelle. London: Edward Arnold.

Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." Journal of Personality and Social Psychology 71:230–244.

Bargh, John A., Katelyn Y. A. McKenna, and Grainne M. Fitzsimons. 2002. "Can You See Me? Activation and Expression of the 'True Self ' on the Internet." Journal of Social Issues 58:33–48.

Barkan, Steven B. and Steven F. Cohn. 1994. "Racial Prejudice and Support for the Death Penalty by Whites." Journal of Research in Crime and Delinquency 31:202–209.

Barrera, Davide and Brent Simpson. 2012. "Much Ado About Deception: Consequences of Deceiving Research Participants in the Social Sciences." Sociological Methods & Research 41:383–413.

Barringer, Felicity. 1993. "Majority in Poll Back Ban on Handguns." The New York Times, June 4, p. A14.

Bartholomew, David J. 2010. "Indirect Measurement." Pp. 455–468 in The SAGE Handbook of Measurement, edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: SAGE.

Baum, Samuel. 1993. “Sources of Demographic Data.” Pp. 3-1–3-50 in Readings in Population Research Methodology. Vol. 1, Basic Tools, edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United Nations Population Fund.

Baumann, Jeannie. 2017. “Proposed Delay to Research Rule Still Up in the Air.” From Health Care on Bloomberg Law, October 18. Retrieved from https://www.bna.com/proposed-delay-research-n73014471089/

Baumrind, Diana. 1964. “Some Thoughts on Ethics of Research: After Reading Milgram’s ‘Behavioral Study of Obedience.’” American Psychologist 19:421–423.

Baumrind, Diana. 1985. “Research Using Intentional Deception: Ethical Issues Revisited.” American Psychologist 40:165–174.

Bean, Lee L., Geraldine P. Mineau, and Douglas L. Anderton. 1990. Fertility Change on the American Frontier: Adaptation and Innovation. Berkeley: University of California Press.

Becker, Howard S. 1958. “Problems of Inference and Proof in Participant Observation.” American Sociological Review 23:652–660.

Becker, Howard S. 1963. The Outsiders: Studies in the Sociology of Deviance. New York: Free Press.

Becker, Howard S. 1974. “Photography and Sociology.” Studies in the Anthropology of Visual Communication 1:3–26.

Becker, Howard S. 2007. Writing for Social Scientists, 2nd ed. Chicago: University of Chicago Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. Habits of the Heart: Individualism and Commitment in American Life. New York: Harper & Row.

Belousov, Konstantin, Tom Horlick-Jones, Michael Bloor, Yakov Gilinskiy, Valentin Golbert, Yakov Kostikovsky, Michael Levi, and Dmitri Pentsov. 2007. “Any Port in a Storm: Fieldwork Difficulties in Dangerous and Crisis-Ridden Settings.” Qualitative Research 7:155–175.

Bendix, Reinhard. 1956. Work and Authority in Industry: Ideologies of Management in the Course of Industrialization. Berkeley: University of California Press.

Bendix, Reinhard. 1962. Max Weber: An Intellectual Portrait. Garden City, NY: Doubleday/Anchor.

Benkler, Yochai. 2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom. New Haven, CT: Yale University Press.

Benwell, Bethan and Elizabeth Stokoe. 2006. Discourse and Identity. Edinburgh, Scotland: Edinburgh University Press.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. “Status Characteristics and Social Interaction.” American Sociological Review 37:241–255.

Bergman, Manfred Max. 2008. “Combining Different Types of Data for Quantitative Analysis.” Pp. 585–601 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992. “The Deterrent Effect of Arrest: A Bayesian Analysis of Four Field Experiments.” American Sociological Review 57 (October):698–708.

Berk, Richard A., Kenneth J. Lenihan, and Peter H. Rossi. 1980. “Crime and Poverty: Some Experimental Evidence From Ex-Offenders.” American Sociological Review 45:766–786.

Berkman, Lisa F. and S. Leonard Syme. 1979. “Social Networks, Host Resistance, and Mortality: A Nine-Year Follow-Up Study of Alameda County Residents.” American

Journal of Epidemiology 109(2):186–204.

Berman, Greg and Aubrey Fox. 2009. Lessons From the Battle Over D.A.R.E.: The Complicated Relationship Between Research and Practice. Washington, DC: Center for Court Innovation and Bureau of Justice Assistance, U.S. Department of Justice.

Berry, Brent. 2006. "Friends for Better or for Worse: Interracial Friendship in the United States as Seen Through Wedding Party Photos." Demography 43:491–510.

Binder, Arnold and James W. Meeker. 1993. "Implications of the Failure to Replicate the Minneapolis Experimental Findings." American Sociological Review 58(December):886–888.

Birkeland, Sarah, Erin Murphy-Graham, and Carol Weiss. 2005. "Good Reasons for Ignoring Good Evaluation: The Case of the Drug Abuse Resistance Education (D.A.R.E.) Program." Evaluation and Program Planning 28:247–256.

Black, Donald J. (Ed.). 1984. Toward a General Theory of Social Control. Orlando, FL: Academic Press.

Black, M. C., K. C. Basile, M. J. Breiding, S. G. Smith, M. L. Walters, M. T. Merrick, J. Chen, & M. R. Stevens. 2011. The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Blau, Peter M. 1964. Exchange and Power in Social Life. New York: Wiley.

Bledsoe, Katrina L. and Rodney K. Hopson. 2009. "Conducting Ethical Research and Evaluation in Underserved Communities." Pp. 391–406 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Bloom, Howard S. 2008. "The Core Analytics of Randomized Experiments for Social Research." Pp. 115–133 in The SAGE Handbook of Social Research Methods, edited by

Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Boase, Jeffrey, John B. Horrigan, Barry Wellman, and Lee Rainie. 2006. The Strength of Internet Ties: The Internet and Email Aid Users in Maintaining Their Social Networks and Provide Pathways to Help When People Face Big Decisions. Washington, DC: Pew Internet & American Life Project.

Bogdewic, Stephan P. 1999. "Participant Observation." Pp. 47–70 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Bogue, Donald J. 1969. Principles of Demography. New York: Wiley.

Bogue, Donald J., Eduardo E. Arriaga, and Douglas L. Anderton. 1993. Readings in Population Research Methodology. Vol. 1, Basic Tools. Chicago: Social Development Center, for the United Nations Population Fund.

Bohannon, John. 2013. "Who's Afraid of Peer Review?" Science 6154:60–65.

Bohn, Angela, Christian Buchta, Kurt Hornik, and Patrick Mair. 2014. "Making Friends and Communicating on Facebook: Implications for the Access to Social Capital." Social Networks 37:29–41.

Bollen, Kenneth A., Barbara Entwisle, and Arthur S. Alderson. 1993. "Macrocomparative Research Methods." Annual Review of Sociology 19:321–351.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." Nature 489:295–298.

Booth, Andrew, Anthea Sutton, and Diana Papaioannou. 2016. Systematic Approaches to a Successful Literature Review, 2nd ed. Thousand Oaks, CA: SAGE.

Booth, Brenda M., Greer Sullivan, Paul Koegel, and Audrey Burnam. 2002. "Vulnerability

Factors for Homelessness Associated With Substance Dependence in a Community Sample of Homeless Adults." American Journal of Drug and Alcohol Abuse 28:429–452.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams. 1995. The Craft of Research. Chicago: University of Chicago Press.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams. 2008. The Craft of Research, 3rd ed. Chicago: University of Chicago Press.

Borg, Marian J. 1997. "The Southern Subculture of Punitiveness? Regional Variation in Support for Capital Punishment." Journal of Research in Crime and Delinquency 34:25–45.

Borg, Marian J. 1998. "Vicarious Homicide Victimization and Support for Capital Punishment: A Test of Black's Theory of Law." Criminology 36:537–567.

Boruch, Robert F. 1997. Randomized Experiments for Planning and Evaluation: A Practical Guide. Thousand Oaks, CA: SAGE.

Bos, Johannes, Aletha Huston, Robert Granger, Greg Duncan, Thomas Brock, and Vonnie McLoyd. 1999. New Hope for People With Low Incomes: Two-Year Results of a Program to Reduce Poverty and Reform Welfare. MRDC Report. Retrieved from www.mdrc.org/publication/new-hope-people-low-incomes

Bosk, Charles L. and Raymond G. De Vries. 2004. "Bureaucracies of Mass Deception: Institutional Review Boards and the Ethics of Ethnographic Research." Annals of the American Academy of Political and Social Science 595:249–263.

Bourgois, Philippe, Mark Lettiere, and James Quesada. 1997. "Social Misery and the Sanctions of Substance Abuse: Confronting HIV Risk Among Homeless Heroin Addicts in San Francisco." Social Problems 44:155–173.

Bray, Hiawatha. 1999. "Plugging in to the Electronic Campus." The Boston Globe

Magazine, April 11, pp. 20–30.

Brenner, Philip S. 2012. “Investigating the Effect of Bias in Survey Measures of Church Attendance.” Sociology of Religion 73:361–383.

Brewer, John and Albert Hunter. 1989. Multimethod Research: A Synthesis of Styles. Newbury Park, CA: SAGE.

Bridges, George S. and Joseph G. Weis. 1989. “Measuring Violent Behavior: Effects of Study Design on Reported Correlates of Violence.” Pp. 14–34 in Violent Crime, Violent Criminals, edited by Neil Alan Weiner and Marvin E. Wolfgang. Newbury Park, CA: SAGE.

Brooks, Clem and Jeff Manza. 2006. “Social Policy Responsiveness in Developed Democracies.” American Sociological Review 71:474–494.

Broskoske, Steve. 2005. “How to Prevent Paper Recycling.” The Teaching Professor 19:1, 4.

Brown, Bruce L. and Dawson Hedges. 2009. “Use and Misuse of Quantitative Methods: Data Collection, Calculation, and Presentation.” Pp. 373–386 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Brown, Judith Belle. 1999. “The Use of Focus Groups in Clinical Research.” Pp. 109–124 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli, and Suzanne L. Wenzel. 2013. “Monogamy on the Street: A Mixed Methods Study of Homeless Men.” Journal of Mixed Methods Research 7:328–346.

Bruni, Frank. 2002. “Persistent Drop in Fertility Reshapes Europe’s Future.” The New York Times, December 26, pp. A1, A10.

Bureau of Justice Statistics. 2011. Key Facts at a Glance: Drug Arrests by Age, 1970–2007. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice. Retrieved from http://bjs.ojp.usdoj.gov/content/glance/drug.cfm

Burke, Garance. 2000. “Mexico’s Census Battles Perception of Corruption.” The Boston Globe, March 2, p. A2.

Burt, Martha R. 1996. “Homelessness: Definitions and Counts.” Pp. 15–23 in Homelessness in America, edited by Jim Baumohl. Phoenix, AZ: Oryx.

Bushman, Brad J., Roy F. Baumeister, and Angela D. Stack. 1999. “Catharsis, Aggression, and Persuasive Influence: Self-Fulfilling or Self-Defeating Prophecies?” Journal of Personality and Social Psychology 76:367–376.

Butler, Declan. 2013. “When Google Got Flu Wrong.” Nature 494(February 14):155–156.

Butler, Dore and Florence Geis. 1990. “Nonverbal Affect Responses to Male and Female Leaders: Implications for Leadership Evaluations.” Journal of Personality and Social Psychology 58(January):48–59.

Buttel, Frederick H. 2000. “World Society, the Nation-State, and Environmental Protection: Comment on Frank, Hironaka, and Schofer.” American Sociological Review 65:117–121.

Buzawa, Eve S. and Carl G. Buzawa (Eds.). 1996. Do Arrests and Restraining Orders Work? Thousand Oaks, CA: SAGE.

Cain, Carol. 1991. “Personal Stories: Identity Acquisition and Self-Understanding in Alcoholics Anonymous.” Ethos 19: 210–253.

Calvey, David. 2017. Covert Research: The Art, Politics and Ethics of Undercover Fieldwork. Thousand Oaks, CA: SAGE.

Campbell, Colin and Jonathan Horowitz. 2016. “Does College Influence Sociopolitical Attitudes?” Sociology of Education 89(1):40–58.

Campbell, Donald T. and Donald W. Fiske. 1959. “Convergent and Discriminant Validation by the Multitrait–Multimethod Matrix.” Psychological Bulletin 56:81–105.

Campbell, Donald T. and M. Jean Russo. 1999. Social Experimentation. Thousand Oaks, CA: SAGE.

Campbell, Donald T. and Julian C. Stanley. 1966. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Campbell, Richard T. 1992. “Longitudinal Research.” Pp. 1146–1158 in Encyclopedia of Sociology, edited by Edgar F. Borgatta and Marie L. Borgatta. New York: Macmillan.

Campbell, Wilson. 2002. “A Statement From the Governmental Accounting Standards Board and Performance Measurement Staff.” American Society for Public Administration. Retrieved from www.aspanet.org/cap/forum_statement.html#top

Cannell, C. F., L. Oksenberg, G. Kalton, K. Bischoping, and Floyd J. Fowler. 1989. New Techniques for Pretesting Survey Questions. Final Report. Grant No. HS05616. National Center for Health Service Research, Health Care Technology Assessment. Retrieved from www.psc.isr.umich.edu/dis/infoserv/isrpub/pdf/NewTechniquesPretestingSurveyQuestio

Caplan, Brina, Russell K. Schutt, Winston M. Turner, Stephen M. Goldfinger, and Larry J. Seidman. 2006. “Change in Neurocognition by Housing Type and Substance Abuse Among Formerly Homeless Seriously Mentally Ill Individuals.” Schizophrenia Research 83:77–86.

Carmines, Edward G. and Richard A. Zeller. 1979. Reliability and Validity Assessment, no. 17, Quantitative Applications in the Social Sciences. Beverly Hills, CA: SAGE.

Catalano, Shannan. 2012. Intimate Partner Violence, 1993–2010. Washington, DC:

Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Cava, Anita, Reid Cushman, and Kenneth Goodman. 2007. "HIPAA and Human Subjects Research." Collaborative Institutional Training Initiative. Retrieved from https://www.citiprogram.org/members/learners

Cave, Emma and Soren Holm. 2003. "Milgram and Tuskegee—Paradigm Research Projects in Bioethics." Health Care Analysis 11:27–40.

Ceglowski, Deborah. 2002. "Research as Relationship." Pp. 5–27 in The Qualitative Inquiry Reader, edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Center for Survey Research, University of Massachusetts at Boston. 1987. "Methodology: Designing Good Survey Questions." Newsletter, April, p. 3.

Centers for Disease Control and Prevention (CDC). 2009. "The Tuskegee Timeline." Atlanta, GA: National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention. Retrieved from www.cdc.gov/tuskegee/timeline.htm

Chadwick, Gary L. 2017. "Final Rule Material: New and Revised Definitions." CITI Program. Retrieved from https://about.citiprogram.org/wp-content/uploads/2017/07/Final-Rule-Material-New-and-Revised-Definitions.pdf

Chase-Dunn, Christopher and Thomas D. Hall. 1993. "Comparing World-Systems: Concepts and Working Hypotheses." Social Forces 71:851–886.

Chen, Huey-Tsyh. 1990. Theory-Driven Evaluations. Newbury Park, CA: SAGE.

Chen, Huey-Tsyh and Peter H. Rossi. 1987. "The Theory-Driven Approach to Validity." Evaluation and Program Planning 10:95–103.

Christian, Leah, Scott Keeter, Kristen Purcell, and Aaron Smith. 2010. Assessing the Cell

Phone Challenge to Survey Research in 2010. Washington, DC: Pew Research Center for the People & the Press and Pew Internet & American Life Project.

Church, Allan H. 1993. “Estimating the Effects of Incentives on Mail Survey Response Rates: A Meta-Analysis.” Public Opinion Quarterly 57:62–79.

Church, A. Timothy. 2010. “Measurement Issues in Cross-Cultural Research.” Pp. 151–175 in The SAGE Handbook of Measurement, edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: SAGE.

CIRCLE. 2013. “Fact Sheet: Voter Turnout Among Young Women and Men in the 2012 Presidential Election.” Medford, MA: The Center for Information & Research on Civil Learning & Engagement, Jonathan M. Tisch College of Citizenship and Public Service, Tufts University. Retrieved from www.civicyouth.org/wp-content/uploads/2013/05/fs_gender_13_final.pdf

CITI Program. 2017. Final Rule Resources. Retrieved from https://about.citiprogram.org/en/final-rule-resources/

Clement, Scott and Peyton M. Craighill. 2015. Gallup Isn’t Doing Any Horserace Polling in 2016. Here’s Why. The Washington Post, October 7. https://www.washingtonpost.com/news/the-fix/wp/2015/10/07/gallup-isnt-doing-any-horserace-polling-in-2016-heres-why/?utm_term=.e1cae30e7250

Clark, Herbert H. and Michael F. Schober. 1994. “Asking Questions and Influencing Answers.” Pp. 15–48 in Questions About Questions: Inquiries Into the Cognitive Bases of Surveys, edited by Judith M. Tanur. New York: Russell Sage Foundation.

Coffey, Amanda and Paul Atkinson. 1996. Making Sense of Qualitative Data: Complementary Research Strategies. Thousand Oaks, CA: SAGE.

Cohen, Alison K. and Benjamin W. Chaffee. 2012. “The Relationship Between Adolescents’ Civic Knowledge, Civic Attitude, and Civic Behavior and Their Self-Reported Future Likelihood of Voting.” Education, Citizenship and Social Justice 8(1):43–57.

Cohen, Sheldon, Robin Mermelstein, Tom Kamarck, and Harry M. Hoberman. 1985. “Measuring the Functional Components of Social Support.” Pp. 73–94 in Social Support: Theory, Research and Applications, edited by Irwin G. Sarason and Barbara R. Sarason. The Hague, The Netherlands: Martinus Nijhoff.

Cohen, Susan G. and Gerald E. Ledford Jr. 1994. “The Effectiveness of Self-Managing Teams: A Quasi-Experiment.” Human Relations 47:13–43.

Cohn, Nate. 2017. “Election Review: Why Crucial State Polls Turned Out to Be Wrong.” The New York Times, June 1:A12.

Coleman, James S. 1990. Foundations of Social Theory. Cambridge, MA: Harvard University Press.

Coleman, James S. and Thomas Hoffer. 1987. Public and Private High Schools: The Impact of Communities. New York: Basic Books.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. High School Achievement: Public, Catholic, and Private Schools Compared. New York: Basic Books.

College Alcohol Study. 2008. “College Alcohol Study News.” Harvard School of Public Health. Retrieved from www.hsph.harvard.edu/cas/Home.html

Collins, Randall. 1994. Four Sociological Traditions. New York: Oxford University Press.

Connolly, Francis J. and Charley Manning. 2001. “What ‘Push Polling’ Is and What It Isn’t.” The Boston Globe, August 16, p. A21.

Connor, Jennie, Andrew Gray, and Kypros Kypri. 2010. “Drinking History, Current Drinking and Problematic Sexual Experiences Among University Students.” Australian and New Zealand Journal of Public Health 34:487–494.

Connors, Gerard J. and Robert J. Volk. 2004. “Self-Report Screening for Alcohol Problems Among Adults.” Pp. 21–35 in Assessing Alcohol Problems: A Guide for Clinicians and

Researchers, 2nd ed., edited by John P. Allen and Veronica B. Wilson. NIH Publication No. 03 -3745, revised 2003. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health. Retrieved October 6, 2005, from https://pubs.niaaa.nih.gov/publications/AssessingAlcohol/index.htm

Converse, Jean M. 1984. “Attitude Measurement in Psychology and Sociology: The Early Years.” Pp. 3–40 in Surveying Subjective Phenomena, vol. 2, edited by Charles F. Turner and Elizabeth Martin. New York: Russell Sage Foundation.

Cook, Gareth. 2005. “Face It: Appearances Sway Voters.” The Boston Globe, June 14, pp. E1, E4.

Cook, Thomas D. and Donald T. Campbell. 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Cook, Thomas D. and Vivian C. Wong. 2008. “Better Quasi-Experimental Practice.” Pp. 134–165 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Cooper, Harris and Larry V. Hedges. 1994. “Research Synthesis as a Scientific Enterprise.” Pp. 3–14 in The Handbook of Research Synthesis, edited by Harris Cooper and Larry V. Hedges. New York: Russell Sage Foundation.

Cooper, Kathleen B. and Nancy J. Victory. 2002. “Foreword.” P. iv in A Nation Online: How Americans Are Expanding Their Use of the Internet. Washington, DC: National Telecommunications and Information Administration, Economics and Statistics Administration, U.S. Department of Commerce.

Core Institute. 1994. Core Alcohol and Drug Survey: Long Form. Carbondale, IL: FIPSE Core Analysis Grantee Group, Core Institute, Student Health Programs, Southern Illinois University.

Correll, Joshua, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink. 2002. “The Police Officer’s Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals.” Journal of Personality and Social Psychology 83:1314–1329.

Corse, Sara J., Nancy B. Hirschinger, and David Zanis. 1995. “The Use of the Addiction Severity Index With People With Severe Mental Illness.” Psychiatric Rehabilitation Journal 19(1):9–18.

Costner, Herbert L. 1989. “The Validity of Conclusions in Evaluation Research: A Further Development of Chen and Rossi’s Theory-Driven Approach.” Evaluation and Program Planning 12:345–353.

Couper, Mick P. 2000. “Web Surveys: A Review of Issues and Approaches.” Public Opinion Quarterly 64:464–494.

Couper, Mick P., Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls II, and James M. O’Reilly (Eds.). 1998. Computer-Assisted Survey Information Collection. New York: Wiley.

Couper, Mick P. and Peter V. Miller. 2008. “Web Survey Methods: Introduction.” Public Opinion Quarterly 72:831–835.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. “Web Survey Design and Administration.” Public Opinion Quarterly 65:230–253.

Cowan, Gloria. 2002. “Content Analysis of Visual Materials.” Pp. 345–368 in Handbook for Conducting Research on Human Sexuality, edited by Michael W. Wiederman and Bernard E. Whitley Jr. London: Lawrence Erlbaum.

Cress, Daniel M. and David A. Snow. 2000. “The Outcomes of Homeless Mobilization: The Influence of Organization, Disruption, Political Mediation, and Framing.” American Journal of Sociology 4:1063–1104.

Creswell, John W. and Vicki L. Plano Clark. 2011. Designing and Conducting Mixed Methods Research, 2nd ed. Thousand Oaks, CA: SAGE.

Cuba, Lee J. 2002. A Short Guide to Writing About Social Science, 4th ed. New York: Addison-Wesley.

Cullen, Kevin. 2014. "The Gunmen, the Shadows, and the Damage Done." Boston Sunday Globe, July 6, pp. A1, A6, A7.

Currie, Janet and Duncan Thomas. 1995. "Does Head Start Make a Difference?" American Economic Review 85:341–365.

Czopp, Alexander M., Margo J. Monteith, and Aimee Y. Mark. 2006. "Standing Up for a Change: Reducing Bias Through Interpersonal Confrontation." Journal of Personality and Social Psychology 90:784–803.

Dabney, Dean A., Laura Dugan, Volkan Topalli, and Richard C. Hollinger. 2006. "The Impact of Implicit Stereotyping on Offender Profiling: Unexpected Results From an Observational Study of Shoplifting." Criminal Justice and Behavior 33:646–675.

Dahlberg, Britt, Marsha N. Wittink, and Joseph J. Gallo. 2010. "Funding and Publishing Integrated Studies: Writing Effective Mixed Methods Manuscripts and Grant Proposals." Pp. 775–802 in SAGE Handbook of Mixed Methods in Social & Behavioral Research, 2nd ed., edited by Abbas Tashakkori and Charles Teddlie. Thousand Oaks, CA: SAGE.

Dale, Angela, Jo Wathan, and Vanessa Higgins. 2008. "Secondary Analysis of Quantitative Data Sources." Pp. 520–535 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

D'Amico, Elizabeth J. and Kim Fromme. 2002. "Brief Prevention for Adolescent Risk-Taking Behavior." Addiction 97:563–574.

Dannefer, W. Dale and Russell K. Schutt. 1982. "Race and Juvenile Justice Processing in Court and Police Agencies." American Journal of Sociology 87(March):1113–1132.

D.A.R.E. 2008. The "New" D.A.R.E. Program. Retrieved from www.dare.com/newdare.asp

D.A.R.E. 2014. “Drug Abuse Resistance Education.” Retrieved from www.dare.org/starting-a-dare-program

Davis, James A. 1985. The Logic of Causal Order. Sage University Paper Series on Quantitative Applications in the Social Sciences, series No. 07–055. Beverly Hills, CA: SAGE.

Davis, James A. and Tom W. Smith. 1992. The NORC General Social Survey: A User’s Guide. Newbury Park, CA: SAGE.

Davis, Ophera A. and Marie Land. 2007. “Southern Women Survivors Speak About Hurricane Katrina, the Children and What Needs to Happen Next.” Race, Gender & Class 14:69–86.

Davis, R. C., D. Weisburd, and B. Taylor. 2008. “Effects of Second Responder Programs on Repeat Incidents of Family Abuse.” Campbell Systematic Reviews, 15.

Dawes, Robyn. 1995. “How Do You Formulate a Testable Exciting Hypothesis?” Pp. 93–96 in How to Write a Successful Research Grant Application: A Guide for Social and Behavioral Scientists, edited by Willo Pequegnat and Ellen Stover. New York: Plenum Press.

de Leeuw, Edith. 2008. “Self-Administered Questionnaires and Standardized Interviews.” Pp. 311–327 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

de Vaus, David. 2008. “Comparative and Cross-National Designs.” Pp. 248–264 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Decker, Scott H. and Barrik Van Winkle. 1996. Life in the Gang: Family, Friends, and Violence. Cambridge: Cambridge University Press.

Deegan, Allison. 2012. “Stranger in a Strange Land: The Challenges and Benefits of Online

Interviews in the Social Networking Space.” Pp. 69–99 in Cases in Online Interview Research, edited by Janet Salmons. Thousand Oaks, CA: SAGE.

Demos, John. 1998. “History Beyond Data Bits.” The New York Times, December 30, p. A23.

Dentler, Robert A. 2002. Practicing Sociology: Selected Fields. Westport, CT: Praeger.

Denzin, Norman K. 2002. “The Interpretive Process.” Pp. 349–368 in The Qualitative Researcher’s Companion, edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: SAGE.

Denzin, Norman and Yvonna S. Lincoln. 2000. “Introduction: The Discipline and Practice of Qualitative Research.” Pp. 1–28 in The Handbook of Qualitative Research, 2nd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Denzin, Norman K. and Yvonna S. Lincoln. 2008. Strategies of Qualitative Inquiry, 3rd ed. Thousand Oaks, CA: SAGE.

DeParle, Jason. 1999. “Project to Rescue Needy Stumbles Against the Persistence of Poverty.” The New York Times, May 15, pp. A1, A10.

Department of Health and Human Services (DHHS). 2009. Code of Federal Regulations, Title 45: Public Welfare. Department of Health and Human Services, Part 46, Protection of Human Subjects. Retrieved from hhs.gov/ohrp/policy/ohrpregulations.pdf

Department of Health, Education, and Welfare. 1979. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Washington, DC: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Office of the Secretary, Department of Health, Education, and Welfare. Retrieved from www.hss.gov/ohrp/humansubjects/guidance/belmont.htm

DeSilver, Drew. 2016. “Turnout Was High in the 2016 Primary Season, But Just Short of

2008 Record." Pew Research Center, June 10. http://www.pewresearch.org/fact-tank/2016/06/10/turnout-was-high-in-the-2016-primary-season-but-just-short-of-2008-record/

Desjardins, Jeff. 2017. "What Happens in an Internet Minute in 2017?" Retrieved from http://www.visualcapitalist.com/happens-internet-minute-2017/

Desmond, Matthew, Andrew Papachristos, and David Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." American Sociological Review 81(5):857–876.

DeVellis, Robert F. 2017. Scale Development: Theory and Applications, 4th ed. Thousand Oaks, CA: SAGE.

Dewan, Shaila K. 2004a. "As Murders Fall, New Tactics Are Tried Against Remainder." The New York Times, December 31, pp. A24–A25.

Dewan, Shaila K. 2004b. "New York's Gospel of Policing by Data Spreads Across U.S." The New York Times, April 26, pp. A1, C16.

Diamond, Timothy. 1992. Making Gray Gold: Narratives of Nursing Home Care. Chicago: University of Chicago Press.

Dickson-Swift, Virginia, Erica L. James, Sandra Kippen, and Pranee Liamputtong. 2008. "Risk to Researchers in Qualitative Research on Sensitive Topics: Issues and Strategies." Qualitative Health Research 18:133–144.

DiClemente, C. C., J. P. Carbonari, R. P. G. Montgomery, and S. O. Hughes. 1994. "The Alcohol Abstinence Self-Efficacy Scale." Journal of Studies on Alcohol 55:141–148.

Dill, Karen E. and Kathryn P. Thill. 2007. "Video Game Characters and the Socialization of Gender Roles: Young People's Perceptions Mirror Sexist Media Depictions." Sex Roles 57:851–864.

Dillman, Don A. 1978. Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Dillman, Don A. 1982. “Mail and Other Self-Administered Questionnaires.” Chapter 12 in Handbook of Survey Research, edited by Peter Rossi, James Wright, and Andy Anderson. New York: Academic Press. As reprinted on pp. 637–638 in Delbert C. Miller, 1991. Handbook of Research Design and Social Measurement, 5th ed. Beverly Hills, CA: SAGE.

Dillman, Don A. 2000. Mail and Internet Surveys: The Tailored Design Method, 2nd ed. New York: Wiley.

Dillman, Don A. 2007. Mail and Internet Surveys: The Tailored Design Method, 2nd ed. Updated With New Internet, Visual, and Mixed-Mode Guide. Hoboken, NJ: Wiley.

Dillman, Don A., James A. Christenson, Edwin H. Carpenter, and Ralph M. Brooks. 1974. “Increasing Mail Questionnaire Response: A Four-State Comparison.” American Sociological Review 39(October):744–756.

Dillman, Don A. and Leah Melani Christian. 2005. “Survey Mode as a Source of Instability in Responses Across Surveys.” Field Methods 17:30–52.

Doucet, Andrea and Natasha Mauthner. 2008. “Qualitative Interviewing and Feminist Research.” Pp. 328–343 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Douglas, Jack D. 1985. Creative Interviewing. Beverly Hills, CA: SAGE.

Drake, Robert E., Gregory J. McHugo, and Jeremy C. Biesanz. 1995. “The Test-Retest Reliability of Standardized Instruments Among Homeless Persons With Substance Use Disorders.” Journal of Studies on Alcohol 56(2):161–167.

Drew, Paul. 2005. “Conversation Analysis.” Pp. 71–102 in Handbook of Language and Social Interaction, edited by Kristine L. Fitch and Robert E. Sanders. Mahwah, NJ:

Lawrence Erlbaum.

Duckworth, Kenneth, John H. Halpern, Russell K. Schutt, and Christopher Gillespie. 2003. “Use of Schizophrenia as a Metaphor in U.S. Newspapers.” Psychiatric Services 54:1402–1404.

Duncombe, Jean and Julie Jessop. 2002. “‘Doing Rapport’ and the Ethics of ‘Faking Friendship.’” Pp. 107–122 in Ethics in Qualitative Research, edited by Melanie Mauthner, Maxine Birch, Julie Jessop, and Tina Miller. Thousand Oaks, CA: SAGE.

Duneier, Mitchell. 1999. Sidewalk. New York: Farrar, Strauss, and Giroux.

Duneier, Mitchell. 2016. Ghetto: the Invention of a Place, the History of an Idea. New York: Farrar, Strauss, and Giroux.

Durkheim, Émile. 1951. Suicide. New York: Free Press.

Durkheim, Émile. 1966. Suicide: A Study in Sociology. Translated by John A. Spaulding and George Simpson. New York: Free Press.

Durkheim, Émile. 1893/2014. The Division of Labor in Society. Edited and with an Introduction by Steven Lukes. Translation by W. D. Halls. New York: Free Press.

Dykema, Jennifer and Nora Cate Schaeffer. 2000. “Events, Instruments, and Reporting Errors.” American Sociological Review 65:619–629.

Eckholm, Erik. 2006. “Report on Impact of Federal Benefits on Curbing Poverty Reignites a Debate.” The New York Times, February 18, p. A8.

Edmonds, W. Alex and Thomas D. Kennedy. 2016. An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods, 2nd ed. Thousand Oaks, CA: SAGE.

Egan, Patrick. 2012. "The Declining Culture of Guns and Violence in the United States." The Monkey Cage Blog, posted July 21. Retrieved from http://themonkeycage.org/2012/07/21/the-declining-culture-of-guns-and-violence-in-the-united-states/

Elder, Keith, Sudha Xirasagar, Nancy Miller, Shelly Ann Bowen, Saundra Glover, and Crystal Piper. 2007. "African Americans' Decisions Not to Evacuate New Orleans Before Hurricane Katrina: A Qualitative Study." American Journal of Public Health 97(Suppl. 1):S124–S129.

Elliott, Jane, Janet Holland, and Rachel Thomson. 2008. "Longitudinal and Panel Studies." Pp. 228–248 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Ellis, Carolyn. 1995. "Emotional and Ethical Quagmires in Returning to the Field." Journal of Contemporary Ethnography 24:68–98.

Ember, Carol R. and Melvin Ember. 2011. A Basic Guide to Cross-Cultural Research. New Haven, CT: Human Relations Area Files, Yale University. Retrieved from www.yale.edu/hraf/guides.htm

Emmel, Nick and Andrew Clark. 2011. "Learning to Use Visual Methodologies in Our Research: A Dialogue Between Two Researchers." Forum: Qualitative Social Research 12(1): Article 36, http://nbn-resolving.de/urn:nbn:de:0114-fqs1101360

Emerson, Robert M. (Ed.). 1983. Contemporary Field Research. Prospect Heights, IL: Waveland.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. Writing Ethnographic Fieldnotes. Chicago: University of Chicago Press.

Emmison, Michael, Philip Smith, and Margery Mayall. 2012. Researching the Visual, 2nd ed. Thousand Oaks, CA: SAGE.

Ennett, Susan T., Karl E. Bauman, Andrea Hussong, Robert Faris, Vangie A. Foshee, and Li Cai. 2006. "The Peer Context of Adolescent Substance Use: Findings From Social Network Analysis." Journal of Research on Adolescence 16:159–186.

Ennett, Susan T., Nancy S. Tobler, Christopher L. Ringwalt, and Robert L. Flewelling. 1994. "How Effective Is Drug Abuse Resistance Education? A Meta-Analysis of Project DARE Outcome Evaluations." American Journal of Public Health 84(9):1394–1401.

Erikson, Kai T. 1966. Wayward Puritans: A Study in the Sociology of Deviance. New York: Wiley.

Erikson, Kai T. 1967. "A Comment on Disguised Observation in Sociology." Social Problems 12: 366–373.

Ettelt, Stefanie, Nicholas Mays, and Ellen Nolte. 2013. "Policy–Research Linkage: What We Have Learned From Providing a Rapid Response Facility for International Healthcare Comparisons to the Department of Health in England." Evidence & Policy 9:245–254.

Eurostat. 2003. ECHP UDB Manual: European Community Household Panel Longitudinal Users' Database. Brussels, Belgium: European Commission.

Exum, M. Lyn. 2002. "The Application and Robustness of the Rational Choice Perspective in the Study of Intoxicated and Angry Intentions to Aggress." Criminology 40:933–966.

Fallon, Kathleen M., Liam Swiss, and Jocelyn Viterna. 2012. "Resolving the Democracy Paradox: Democratization and Women's Legislative Representation in Developing Nations, 1975 to 2009." American Sociological Review 77(3):380–408.

Fawcett, Barbara and Rosalie Pockett. 2015. Turning Ideas Into Research: Theory, Design and Practice. Thousand Oaks, CA: SAGE.

Fears, Darryl. 2002. "For Latinos in U.S., Race Not Just Black or White." The Boston Globe, December 30, p. A3.

Federal Bureau of Investigation. n.d. "NIBRS Overview." Retrieved from https://ucr.fbi.gov/nibrs-overview

Federal Register. 2017. Federal Policy for the Protection of Human Subjects, 82(12, Part IX), 49 CFR Part 11, January 19:7149–7274. Retrieved from https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/html/2017-01058.htm

Fedyuk, Olena. 2012. "Images of Transnational Motherhood: The Role of Photographs in Measuring Time and Maintaining Connections Between Ukraine and Italy." Journal of Ethnic and Migration Studies 38:279–300.

Fenno, Richard F. Jr. 1978. Home Style: House Members in Their Districts. Boston: Little, Brown.

Fenton, Steve. 1996. "Counting Ethnicity: Social Groups and Official Categories." Pp. 143–165 in Interpreting Official Statistics, edited by Ruth Levitas and Will Guy. New York: Routledge.

Ferguson, Kristin M., Kimberly Bender, Sanna J. Thompson, Elaine M. Maccio, and David Pollio. 2012. "Employment Status and Income Generation Among Homeless Young Adults: Results From a Five-City, Mixed-Methods Study." Youth & Society 44:385–407.

File, Thom. 2013. "Computer and Internet Use in the United States." Pp. 20–569 in Current Population Survey Reports. Washington, DC: U.S. Census Bureau.

Fink, Arlene. 2005. Conducting Research Literature Reviews: From the Internet to Paper, 2nd ed. Thousand Oaks, CA: SAGE.

Finkel, Eli J., Paul W. Eastwick, and Harry T. Reis. 2017. "Replicability and Other Features of a High-Quality Science: Toward a Balanced and Empirical Approach." Journal of Personality and Social Psychology 113:244–253.

Firebaugh, Glenn. 2008. Seven Rules for Social Research. Princeton, NJ: Princeton

University Press.

Fischer, Claude. 2009. “The 2004 GSS Finding of Shrunken Social Networks: An Artifact?” American Sociological Review 74(4):657–669.

Fischer, Constance T. and Frederick J. Wertz. 2002. “Empirical Phenomenological Analyses of Being Criminally Victimized.” Pp. 275–304 in The Qualitative Researcher’s Companion, edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: SAGE.

Fisher, Celia B. and Andrea E. Anushko. 2008. “Research Ethics in Social Science.” Pp. 94–109 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Fligstein, Neil, Orestes P. Hastings, and Adam Goldstein. 2017. “Keeping Up With the Joneses: How Households Fared in the Era of High Income Inequality and the Housing Price Bubble, 1999–2007. Socius: Sociological Research for a Dynamic World 3:1–15.

Forero, Juan. 2000a. “Census Takers Say Supervisors Fostered Filing of False Data.” The New York Times, July 28, p. A21.

Forero, Juan. 2000b. “Census Takers Top ‘90 Efforts in New York City, With More to Go.” The New York Times, June 12, p. A29.

Fowler, Floyd J. 1988. Survey Research Methods, Rev. ed. Newbury Park, CA: SAGE.

Fowler, Floyd J. 1995. Improving Survey Questions: Design and Evaluation. Thousand Oaks, CA: SAGE.

Fox, Chris, Robert Grimm, and Rute Caldeira. 2016. An Introduction to Evaluation. Thousand Oaks, CA: SAGE.

Fox, Nick and Chris Roberts. 1999. “GPs in Cyberspace: The Sociology of a ‘Virtual Community.’” The Sociological Review 47:643–669.

Frank, David John, Ann Hironaka, and Evan Schofer. 2000. “The Nation-State and the Natural Environment Over the Twentieth Century.” American Sociological Review 65:96–116.

Frankfort-Nachmias, Chava and Anna Leon-Guerrero. 2006. Social Statistics for a Diverse Society, 4th ed. Thousand Oaks, CA: SAGE.

Franklin, Mark N. 1996. “Electoral Participation.” Pp. 216–235 in Comparing Democracies: Elections and Voting in Global Perspective, edited by Lawrence LeDuc, Richard G. Niemi, and Pippa Norris. Thousand Oaks, CA: SAGE.

Freshman, Audrey. 2012. “Financial Disaster as a Risk Factor for Posttraumatic Stress Disorder: Internet Survey of Trauma in Victims of the Madoff Ponzi Scheme.” Health & Social Work 37:39–48.

Frohmann, Lisa. 2005. “The Framing Safety Project: Photographs and Narratives by Battered Women.” Violence Against Women 11:1396–1419.

Gaiser, Ted J. and Anthony E. Schreiner. 2009. A Guide to Conducting Online Research. Thousand Oaks, CA: SAGE.

Gall, Carlotta. 2003. “Armed With Pencils, Army of Census Workers Fans Out Into Afghan Outback.” The New York Times, July 15, p. A4.

Gallup. 2011. Election Polls—Accuracy Record in Presidential Elections. Retrieved from http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print

Gans, Curtis. 2008. “2008 Primary Turnout Falls Just Short of Record Nationally, Breaks Records in Most States.” AU News, May 19. Washington, DC: Center for the Study of the American Electorate, American University. Retrieved from www.american.edu/media

Garfinkel, Harold. 1967. Studies in Ethnomethodology. Englewood Cliffs, NJ: Prentice Hall.

Geertz, Clifford. 1973. “Thick Description: Toward an Interpretive Theory of Culture.” Pp. 3–30 in The Interpretation of Cultures, edited by Clifford Geertz. New York: Basic Books.

Gerth, Matthias A. and Gabriele Siegert. 2012. “Patterns of Consistence and Constriction: How News Media Frame the Coverage of Direct Democratic Campaigns.” American Behavioral Scientist 56:279–299.

“‘Get Tough’ Youth Programs Are Ineffective, Panel Says.” 2004. The New York Times, October 17, p. 25.

Gil, Paul. 2017. “The Best Search Engines of 2017,” Lifewire, September 11. Retrieved from https://www.lifewire.com/best-search-engines-2483352

Gilchrist, Valerie J. and Robert L. Williams. 1999. “Key Informant Interviews.” Pp. 71–88 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Giles, Jim. 2005. “Internet Encyclopaedias Go Head to Head.” Nature 438:900–901.

Gill, Hannah E. 2004. “Finding a Middle Ground Between Extremes: Notes on Researching Transnational Crime and Violence.” Anthropology Matters Journal 6:1–9.

Gill, Richard T., Nathan Glazer, and Stephan A. Thernstrom. 1992. Our Changing Population. Englewood Cliffs, NJ: Prentice Hall.

Gilligan, Carol. 1988. “Adolescent Development Reconsidered.” Pp. vii–xxxix in Mapping the Moral Domain, edited by Carol Gilligan, Janie Victoria Ward, and Jill McLean Taylor. Cambridge, MA: Harvard University Press.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. “Detecting Influenza Epidemics Using Search Engine Query Data.” Nature 457(February 19):1012–1015.

Glaser, Barney G. and Anselm L. Strauss. 1967. The Discovery of Grounded Theory: Strategies for Qualitative Research. London: Weidenfeld and Nicholson.

Glover, Judith. 1996. “Epistemological and Methodological Considerations in Secondary Analysis.” Pp. 28–38 in Cross-National Research Methods in the Social Sciences, edited by Linda Hantrais and Steen Mangen. New York: Pinter.

Glueck, Sheldon and Eleanor Glueck. 1950. Unraveling Juvenile Delinquency. New York: Commonwealth Fund.

Gobo, Giampietro. 2008. “Re-Conceptualizing Generalization: Old Issues in a New Frame.” Pp. 193–213 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Goertz, Gary. 2006. Social Science Concepts: A User’s Guide. Princeton, NJ: Princeton University Press.

Goffman, Erving. 1961. Asylums: Essays on the Social Situation of Mental Patients and Other Inmates. Garden City, NY: Doubleday.

Goffman, Erving. 1963. Stigma: Notes on the Management of Spoiled Identity. Englewood, NJ: Prentice Hall/Spectrum.

Goffman, Erving. 1979. Gender Advertisements. London: Macmillan.

Gold, Raymond. 1958. “Roles in Sociological Field Observation.” Social Forces 36:217–223.

Golder, Scott A. and Michael W. Macy. 2011. “Diurnal and Seasonal Mood Vary With Work, Sleep, and Daylength Across Diverse Cultures.” Science 333:1878.

Golder, Scott A. and Michael W. Macy. 2014. “Digital Footprints: Opportunities and Challenges for Online Social Research.” Annual Review of Sociology 40:129–152.

Goldfinger, Stephen M. and Russell K. Schutt. 1996. "Comparison of Clinicians' Housing Recommendations and Preferences of Homeless Mentally Ill Persons." Psychiatric Services, 47:413–415.

Goldfinger, Stephen M., Russell K. Schutt, Larry J. Seidman, Winston M. Turner, Walter E. Penk, and George S. Tolomiczenko. 1996. "Self-Report and Observer Measures of Substance Abuse Among Homeless Mentally Ill Persons in the Cross-Section and Over Time." The Journal of Nervous and Mental Disease 184(11):667–672.

Goleman, Daniel. 1993a. "Placebo Effect Is Shown to Be Twice as Powerful as Expected." The New York Times, August 17, p. C3.

Goleman, Daniel. 1993b. "Pollsters Enlist Psychologists in Quest for Unbiased Results." The New York Times, September 7, pp. C1, C11.

Goode, Erich. 2002. "Sexual Involvement and Social Research in a Fat Civil Rights Organization." Qualitative Sociology 25:501–534.

Goodnough, Abby. 2010. "A Wave of Addiction and Crime, With the Medicine Cabinet to Blame." The New York Times, September 22.

Gordon, Raymond. 1992. Basic Interviewing Skills. Itasca, IL: Peacock.

Gough, David, Sandy Oliver, and James Thomas. 2017. An Introduction to Systematic Reviews, 2nd ed. Thousand Oaks, CA: SAGE.

Graduate Program in Applied Sociology. 1990. Handbook for Thesis Writers. Boston: University of Massachusetts.

Grady, John. 1996. "The Scope of Visual Sociology." Visual Sociology 11:10–24.

Grant, Bridget F., Deborah A. Dawson, Frederick S. Stinson, S. Patricia Chou, Mary C. Dufour, and Roger P. Pickering. 2004. "The 12-Month Prevalence and Trends in DSM-IV Alcohol Abuse and Dependence: United States, 1991–1992 and 2001–2002." Drug and Alcohol Dependence 74:223–234.

Greenberg, David, Mark Shroder, and Mathew Onstott. 1999. The Digest of Social Experiments, 2nd ed. Washington, DC: Urban Institute Press.

Grey, Robert J. Jr. 2005. "Jury Service: It's a Privilege." American Bar Association. Retrieved from www.abanet.org/media/releases/opedjuror2.html

Grieco, Elizabeth M., Yesenia D. Acosta, G. Patricia de la Cruz, Christine Gambino, Thomas Gryn, Luke J. Larsen, Edward N. Trevelyan, and Nathan P. Walters. 2012. The Foreign-Born Population in the United States: 2010. Washington, DC: U.S. Census Bureau.

Griffin, Larry J. 1992. "Temporality, Events, and Explanation in Historical Sociology: An Introduction." Sociological Methods & Research 20:403–427.

Griffin, Larry J. 1993. "Narrative, Event-Structure Analysis, and Causal Interpretation in Historical Sociology." American Journal of Sociology 98(March):1094–1133.

Grinnell, Frederick. 1992. The Scientific Attitude, 2nd ed. New York: Guilford.

Grissom, Brandi. 2011. "Proposals Could Make It Harder to Leave Prison." The New York Times, March 12.

Groves, Robert M. 1989. Survey Errors and Survey Costs. New York: Wiley.

Groves, Robert M. and Mick P. Couper. 1998. Nonresponse in Household Interview Surveys. New York: Wiley.

Groves, Robert M. and Robert L. Kahn. 1979. Surveys by Telephone: A National Comparison With Personal Interviews. New York: Academic Press. As adapted in Delbert C. Miller, 1991. Handbook of Research Design and Social Measurement, 5th ed. Newbury Park, CA: SAGE.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Salience Theory of Survey Participation: Description and an Illustration." Public Opinion Quarterly 64:299–308.

Gruenewald, Paul J., Andrew J. Treno, Gail Taff, and Michael Klitzner. 1997. Measuring Community Indicators: A Systems Approach to Drug and Alcohol Problems. Thousand Oaks, CA: SAGE.

Guba, Egon G. and Yvonna S. Lincoln. 1989. Fourth Generation Evaluation. Newbury Park, CA: SAGE.

Guba, Egon G. and Yvonna S. Lincoln. 1994. "Competing Paradigms in Qualitative Research." Pp. 105–117 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Gubrium, Jaber F. and James A. Holstein. 1997. The New Language of Qualitative Method. New York: Oxford University Press.

Gubrium, Jaber F. and James A. Holstein. 2000. "Analyzing Interpretive Practice." Pp. 487–508 in The Handbook of Qualitative Research, 2nd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Guterbock, Thomas M. 2008. Strategies and Standards for Reaching Respondents in an Age of New Technology. Presentation to the Harvard Program on Survey Research Spring Conference, New Technologies and Survey Research. Cambridge, MA: Institute of Quantitative Social Science, Harvard University, May 9.

Hacker, Karen. 2013. Community-Based Participatory Research. Thousand Oaks, CA: SAGE.

Hacker, Karen, Jessica Collins, Leni Gross-Young, Stephanie Almeida, and Noreen Burke. 2008. "Coping With Youth Suicide and Overdose: One Community's Efforts to Investigate, Intervene, and Prevent Suicide Contagion." Crisis 29:86–95.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. “What the Polls Don’t Show: A Closer Look at U.S. Church Attendance.” American Sociological Review 58(6):741–752.

Hafner, Katie. 2004. “For Some, the Blogging Never Stops.” The New York Times, May 27, pp. E1, E7.

Hage, Jerald and Barbara Foley Meeker. 1988. Social Causality. Boston: Unwin Hyman.

Hakim, Catherine. 1982. Secondary Analysis in Social Research: A Guide to Data Sources and Methods With Examples. London: George Allen & Unwin.

Hakimzadeh, Shirin and D’Vera Cohn. 2007. English Usage Among Hispanics in the United States. Washington, DC: Pew Hispanic Center. Retrieved from http://pewhispanic.org/files/reports/82.pdf

Hallinan, Maureen T. 1997. “The Sociological Study of Social Change.” American Sociological Review 62:1–11.

Hammersley, Martyn. 2008. “Assessing Validity in Social Research.” Pp. 42–53 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Hammersley, Martyn and Anna Traianou. 2012. Ethics in Qualitative Research: Controversies and Contexts. Thousand Oaks, CA: SAGE.

Hampton, Keith N. 2003. “Grieving for a Lost Network: Collective Action in a Wired Suburb.” The Information Society 19:417–428.

Hampton, Keith N., Lauren Sessions Goulet, Eun Ja Her, and Lee Rainie. 2009. Social Isolation and New Technology: How the Internet and Mobile Phones Impact Americans’ Social Networks. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/2009/11/04/social-isolation-and-new-technology/

Hampton, Keith N., Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. Social Networking Sites and Our Lives: How People's Trust, Personal Relationships, and Civic and Political Involvement Are Connected to Their Use of Social Networking Sites and Other Technologies. Washington, DC: Pew Internet & American Life Project. Retrieved from http://pewinternet.org/Reports/2011/Technology-and-social-networks.aspx

Hampton, Keith N. and Neeti Gupta. 2008. "Community and Social Interaction in the Wireless City: Wi-Fi Use in Public and Semi-Public Spaces." New Media & Society 10(6):831–850.

Hampton, Keith N. and Barry Wellman. 1999. "Netville On-line and Off-Line: Observing and Surveying a Wired Suburb." American Behavioral Scientist 43:475–492.

Hampton, Keith N. and Barry Wellman. 2000. "Examining Community in the Digital Neighborhood: Early Results From Canada's Wired Suburb." Pp. 475–492 in Digital Cities: Technologies, Experiences, and Future Perspectives, edited by Toru Ishida and Katherine Isbister. Berlin, Germany: Springer-Verlag.

Hampton, Keith N. and Barry Wellman. 2001. "Long Distance Community in the Network Society." American Behavioral Scientist 45:476–495.

Haney, C., C. Banks, and Philip G. Zimbardo. 1973. "Interpersonal Dynamics in a Simulated Prison." International Journal of Criminology and Penology 1:69–97.

Hantrais, Linda and Steen Mangen. 1996. "Method and Management of Cross-National Social Research." Pp. 1–12 in Cross-National Research Methods in the Social Sciences, edited by Linda Hantrais and Steen Mangen. New York: Pinter.

Hard, Stephen F., James M. Conway, and Antonia C. Moran. 2006. "Faculty and College Student Beliefs About the Frequency of Student Academic Misconduct." The Journal of Higher Education 77:1058–1080.

Harding, David J. 2007. "Cultural Context, Sexual Behavior, and Romantic Relationships in Disadvantaged Neighborhoods." American Sociological Review 72:341–364.

Harris, David R. and Jeremiah Joseph Sim. 2002. "Who Is Multiracial? Assessing the Complexity of Lived Race." American Sociological Review 67:614–627.

Harris Poll. 2017. "Rejecting Labels and Sparking an Identity Revolution, Millennials Refuse to be Trumped." Retrieved from http://www.theharrispoll.com/politics/Rejecting-Labels-Identity-Revolution.html.

Hart, Chris. 1998. Doing a Literature Review: Releasing the Social Science Research Imagination. London: SAGE.

Hasin, Deborah S., Frederick S. Stinson, Elizabeth Ogburn, and Bridget F. Grant. 2007. "Prevalence, Correlates, Disability, and Comorbidity of *DSM-IV* Alcohol Abuse and Dependence in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions." Archives of General Psychiatry 64:830–842.

Hawkins, Donald N., Paul R. Amato, and Valarie King. 2007. "Nonresident Father Involvement and Adolescent Well-Being: Father Effects or Child Effects?" American Sociological Review 72: 990–1010.

Healy, Jack. 2017. "For One Family in Houston, an Overwhelming Start to Recovery." The New York Times, September 2. Retrieved from https://www.nytimes.com/interactive/2017/09/01/upshot/cost-of-hurricane-harvey-only-one-storm-comes-close.html

Heath, Christian and Paul Luff. 2008. "Video and the Analysis of Work and Interaction." Pp. 493–505 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Heaton, Janet. 2008. "Secondary Analysis of Qualitative Data." Pp. 506–535 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." Social Problems 44:174–199.

Heckathorn, Douglas D. 2002. “Respondent-Driven Sampling II: Deriving Valid Population Estimates From Chain-Referral Samples of Hidden Populations.” Social Problems 49:11–34.

Heckman, James, Neil Hohmann, and Jeffrey Smith. 2000. “Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment.” The Quarterly Journal of Economics 115:651–694.

Hedström, Peter and Richard Swedberg (Eds.). 1998. Social Mechanisms: An Analytical Approach to Social Theory. Cambridge: Cambridge University Press.

Heise, David R. 2010. Surveying Cultures: Discovering Shared Conceptions and Sentiments. Hoboken, NJ: Wiley.

Herek, Gregory. 1995. “Developing a Theoretical Framework and Rationale for a Research Proposal.” Pp. 85–91 in How to Write a Successful Research Grant Application: A Guide for Social and Behavioral Scientists, edited by Willo Pequegnat and Ellen Stover. New York: Plenum Press.

Heritage, John. n.d. Conversation Analysis. Chapter in E-Source, Social and Behavioral Sciences Research. Washington, DC: Office of Behavioral and Social Sciences Research, National Institutes of Health. Retrieved from www.esourceresearch.org/eSourceBook/ConversationAnalysis/10TranscriptionSymbols/ta

Hesse-Biber, Sharlene. 1989. “Eating Problems and Disorders in a College Population: Are College Women’s Eating Problems a New Phenomenon?” Sex Roles 20:71–89.

Hewson, Claire, Carl Vogel, and Dianna Laurent. 2016. Internet Research Methods, 2nd ed. Thousand Oaks, CA: SAGE.

Hicks, Lorna. 2013. The Regulations—SBE. Miami, FL: Collaborative Institutional Training Initiative at the University of Miami. Retrieved from https://www.citiprogram.org/

The Hippocratic Oath. n.d. Charlottesville: University of Virginia, Historica Medicina. Retrieved from http://exhibits.hsl.virginia.edu/antiqua/hippocrates/

Hirsch, Kathleen. 1989. Songs From the Alley. New York: Doubleday.

Hirschel, David, Eve Buzawa, April Pattavina, and Don Faggiani. 2008. "Domestic Violence and Mandatory Arrest Laws: To What Extent Do They Influence Police Arrest Decisions?" Journal of Criminal Law & Criminology 98(1):255–298.

Ho, D. Y. F. 1996. "Filial Piety and Its Psychological Consequences." Pp. 155–165 in Handbook of Chinese Psychology, edited by M. H. Bond. Hong Kong: Oxford University Press.

Hofling, Charles K., Eveline M. Brotzman, Sarah M. Dalrymple, Nancy M. Graves, and Chester M. Pierce. 1966. "An Experimental Study of Nurse-Physician Relations." Journal of Nervous and Mental Disease 143:171–180.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." Public Opinion Quarterly 60:58–88.

Hollingsworth, T. H. 1972. "The Importance of the Quality of Data in Historical Demography." Pp. 71–86 in Population and Social Change, edited by D. V. Glass and Roger Revelle. London: Edward Arnold.

Holmes, Steven A. 2000. "Stronger Response by Minorities Helps Improve Census Reply Rate." The New York Times, May 4, pp. A1, A22.

Holmes, Steven A. 2001a. "Census Officials Ponder Adjustments Crucial to Redistricting." The New York Times, February 12, p. A17.

Holmes, Steven A. 2001b. "The Confusion Over Who We Are." The New York Times, June 3, p. WK 1.

Horney, Julie, D. Wayne Osgood, and Ineke Haen Marshall. 1995. "Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances." American Sociological Review 60:655–673.

Horowitz, Carol R., Mimsie Robinson, and Sarena Seifer. 2009. "Community-Based Participatory Research From the Margin to the Mainstream: Are Researchers Prepared?" Circulation 119:2633–2642.

Houtzager, Peter P. and Arnab K. Acharya. 2011. "Associations, Active Citizenship, and the Quality of Democracy in Brazil and Mexico." Theory and Society 40:1–36.

Howell, James C. 2003. Preventing and Reducing Juvenile Delinquency: A Comprehensive Framework. Thousand Oaks, CA: SAGE.

Hoyle, Carolyn and Andrew Sanders. 2000. "Police Response to Domestic Violence: From Victim Choice to Victim Empowerment." British Journal of Criminology 40:14–26.

Hrobjartsson, Asbjorn and Peter C. Gotzsche. 2001. "Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo With No Treatment." New England Journal of Medicine 344:1594–1602.

Hu, Winnie. 2014. "Severe Cold Moves New York's Homeless to Seek Help." The New York Times, January 28, p. A19.

Huberman, A. Michael and Matthew B. Miles. 1994. "Data Management and Analysis Methods." Pp. 428–444 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Huff, Darrell. 1954. How to Lie With Statistics. New York: Norton.

Humes, Karen R., Nicholas A. Jones, and Roberto R. Ramirez. 2011. Overview of Race and Hispanic Origin: 2010. C2010BR-02. Washington, DC: U.S. Census Bureau.

Humphrey, Nicholas. 1992. A History of the Mind: Evolution and the Birth of

Consciousness. New York: Simon & Schuster.

Humphreys, Laud. 1970. Tearoom Trade: Impersonal Sex in Public Places. Chicago: Aldine de Gruyter.

Humphries, Courtney. 2011. "Deeply Conflicted: How Can We Insulate Ourselves From Conflicts of Interest? The Most Popular Solution—Disclosing Them—Turns Out Not to Help." Boston Globe, May 15, pp. K1, K3.

Hunt, Morton. 1985. Profiles of Social Research: The Scientific Study of Human Interactions. New York: Russell Sage Foundation.

Huston, Patricia and C. David Naylor. 1996. "Health Services Research: Reporting on Studies Using Secondary Data Sources." Canadian Medical Association Journal 155:1697–1702.

Hyvärinen, Matti. 2008. "Analyzing Narratives and Story-Telling." Pp. 447–460 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Iezzoni, Lisa I. 1997. "How Much Are We Willing to Pay for Information About Quality of Care?" Annals of Internal Medicine 126:391–393.

Innstrand, Siw Tone, Geir Arild Espries, and Reidar Mykletun. 2004. "Job Stress, Burnout and Job Satisfaction: An Intervention Study of Staff Working With People With Intellectual Disabilities." Journal of Applied Research in Intellectual Disabilities 17:119–126.

Internet World Stats. 2017. Internet World Stats: Usage and Population Statistics. Retrieved from http://www.internetworldstats.com/stats.htm

Irvine, Leslie. 1998. "Organizational Ethics and Fieldwork Realities: Negotiating Ethical Boundaries in Codependents Anonymous." Pp. 167–183 in Doing Ethnographic Research: Fieldwork Settings. Thousand Oaks, CA: SAGE.

Israel, Mark. 2014. Research Ethics and Integrity for Social Scientists: Beyond Regulatory Compliance, 2nd ed. Thousand Oaks, CA: SAGE.

Jalbert, Sarah Kuck, William Rhodes, Christopher Flygare, and Michael Kane. 2010. "Testing Probation Outcomes in an Evidence-Based Practice Setting: Reduced Caseload Size and Intensive Supervision Effectiveness." Journal of Offender Rehabilitation, 49:233–253

James, Nalita and Hugh Busher. 2009. Online Interviewing. Thousand Oaks, CA: SAGE.

Jang, Bohyun Joy, Megan E. Patrick, and Megan S. Schuler. 2017. "Substance Use Behaviors and the Timing of Family Formation During Young Adulthood." Journal of Family Issues, Online first.

Jarvis, Helen. 1997. "Housing, Labour Markets and Household Structure: Questioning the Role of Secondary Data Analysis in Sustaining the Polarization Debate." Regional Studies 31:521–531.

Jervis, Robert. 1996. "Counterfactuals, Causation, and Complexity." Pp. 309–316 in Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives, edited by Philip E. Tetlock and Aaron Belkin. Princeton, NJ: Princeton University Press.

Jesnadum, Anick. 2000. "Researchers Fear Privacy Breaches With Online Research." Digital Mass. Retrieved from www.digitalmass.com/news/daily/09/15/researchers.html

Johnson, David and Merry Bullock. 2009. "The Ethics of Data Archiving: Issues From Four Perspectives." Pp. 214–228 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Jones, Stephen R. G. 1992. "Was There a Hawthorne Effect?" American Journal of Sociology 98:451–468.

Kagay, Michael R. with Janet Elder. 1992. "Numbers Are No Problem for Pollsters. Words

Are." The New York Times, October 9, p. E5.

Kale-Lostuvali, Elif. 2007. "Negotiating State Provision: State-Citizen Encounters in the Aftermath of the İzmit Earthquake." The Sociological Quarterly 48:745–767.

Kaplan, Fred. 2002. "NY Continues to See Plunge in Number of Felonies." The Boston Globe, April 15, p. A3.

Kato, Yuki, Catarina Passidomo, and Daina Harvey. 2014. "Political Gardening in Post-Disaster City: Lessons From New Orleans." Urban Studies 51:1833–1849.

Kaufman, Sharon R. 1986. The Ageless Self: Sources of Meaning in Late Life. Madison: University of Wisconsin Press.

Keeter, Scott. 2008. Survey Research and Cell Phones: Is There a Problem? Presentation to the Harvard Program on Survey Research Spring Conference, New Technologies and Survey Research. Cambridge, MA: Institute of Quantitative Social Science, Harvard University, May 9.

Keeter, Scott, Nick Hatley, Courtney Kennedy, and Arnold Lau. 2017. "What Low Response Rates Mean for Telephone Surveys." Pew Research Center, May 15. Retrieved from http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/

Kemmis, Stephen and Robin McTaggart. 2005. "Participatory Action Research: Communicative Action and the Public Sphere." Pp. 559–603 in The SAGE Handbook of Qualitative Research, 3rd ed., edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Kemp, Simon. 2017. "The Global State of the Internet in 20107." Retrieved from https://thenextweb.com/contributors/2017/04/11/current-global-state-internet/

Kennedy, David M., Anthony M. Piehl, and Anne A. Braga. 1996. "Youth Violence in Boston: Gun Markets, Serious Youth Offenders, and a Use-Reduction Strategy." Law and Contemporary Problems 59:147–196.

Kenney, Charles. 1987. “They’ve Got Your Number.” The Boston Globe Magazine, August 30, pp. 12, 46–56, 60.

Kershaw, David and Jerilyn Fair. 1976. The New Jersey Income-Maintenance Experiment, vol. 1. New York: Academic Press.

Kershaw, Sarah. 2000. “In a Black Community, Mistrust of Government Hinders Census.” The New York Times, May 16, p. A20.

Kifner, John. 1994. “Pollster Finds Error on Holocaust Doubts.” The New York Times, May 20, p. A12.

Kilger, Magnus. 2017. “Talent Stories in Youth Sports: Discursively Shared Narratives of Success.” Narrative Inquiry 27:47–65.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. Scientific Inference in Qualitative Research. Princeton, NJ: Princeton University Press.

King, Miriam L. and Diana L. Magnuson. 1995. “Perspectives on Historical U.S. Census Undercounts.” Social Science History 19: 455–466.

King, Nigel and Christine Horrocks. 2010. Interviews in Qualitative Research. Thousand Oaks, CA: SAGE.

Kirk, Andy. 2016. Data Visualisation: A Handbook for Data Driven Design. Thousand Oaks, CA: SAGE.

Kiser, Edgar and Michael Hechter. 1991. “The Role of General Theory in Comparative-Historical Sociology.” American Journal of Sociology 97:1–30.

Kitchener, Karen Strohm and Richard F. Kitchener. 2009. “Social Science Research Ethics:

Historical and Philosophical Issues." Pp. 5–22 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Kitchin, Rob, Tracey P. Lauriault, and Matthew W. Wilson. 2017. Understanding Spatial Media. Thousand Oaks, CA: SAGE.

Klein, Malcolm W. 1971. Street Gangs and Street Workers. Englewood Cliffs, NJ: Prentice Hall.

Klinenberg, Eric. 2002. Heat Wave: A Social Autopsy of Disaster in Chicago. Chicago: University of Chicago Press.

Koegel, Paul. 1987. Ethnographic Perspectives on Homeless and Homeless Mentally Ill Women. Washington, DC: Alcohol, Drug Abuse, and Mental Health Administration, Public Health Service, U.S. Department of Health and Human Services.

Kohn, Alfie. 2008. "Who's Cheating Whom?" The Education Digest 73:4–11.

Kohn, Melvin L. 1987. "Cross-National Research as an Analytic Strategy." American Sociological Review 52:713–731.

Kohut, Andrew. 1988. "Polling: Does More Information Lead to Better Understanding?" The Boston Globe, November 7, p. 25.

Kohut, Andrew, Scott Keeter, Carroll Doherty, Michael Dimock, and Leah Christian. 2012. Assessing the Representativeness of Public Opinion Surveys. Washington, DC: Pew Research Center. Retrieved from www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys

Kolbert, Elizabeth. 1992. "Test-Marketing a President." The New York Times Magazine, August 30, pp. 18–21, 60, 68, 72.

Kollock, Peter and Marc A. Smith. 1999. "Communities in Cyberspace." Pp. 3–25 in

Communities in Cyberspace, edited by Peter Kollock and Marc A. Smith. New York: Routledge.

Korn, James H. 1997. Illusions of Reality: A History of Deception in Social Psychology. Albany: State University of New York Press.

Kotkin, Stephen. 2002. "A World War Among Professors." The New York Times, September 7, Arts pp. 1, 17.

Kozinets, Robert V. 2010. Netnography: Doing Ethnographic Research Online. Thousand Oaks, CA: SAGE.

Kraemer, Helena Chmura and Sue Thiemann. 1987. How Many Subjects? Statistical Power Analysis in Research. Newbury Park, CA: SAGE.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks." Proceedings of the National Academy of Sciences (PNAS), 111:8788–8790.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." Public Opinion Quarterly 72:847–865.

Krogstad, Jens Manuel, Mark Hugo Lopez, and Molly Rohal. 2015. English Proficiency on the Rise Among Latinos: U.S. Born Driving Language Changes. Washington, DC: Pew Research Center. Retrieved from http://www.pewhispanic.org/2015/05/12/english-proficiency-on-the-rise-among-latinos/

Krosnick, Jon A. 1999. "Survey Research." Annual Review of Psychology 50:537–567.

Krosnick, Jon A. 2006. The Handbook of Questionnaire Design: Insights From Social and Cognitive Psychology. Eric M. Mindich Encounters With Authors Symposium at Institute for Quantitative Social Science, Harvard University, Cambridge MA, January 19–21.

Krueger, Richard A. and Mary Anne Casey. 2009. Focus Groups: A Practical Guide for Applied Research, 4th ed. Thousand Oaks, CA: SAGE.

Kubey, Robert. 1990. "Television and the Quality of Family Life." Communication Quarterly 38(Fall):312–324.

Kuhn, Thomas S. 1970. The Structure of Scientific Revolutions, 2nd ed. Chicago: University of Chicago Press.

Kuzel, Anton J. 1999. "Sampling in Qualitative Inquiry." Pp. 33–45 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Kvale, Steinar. 1996. Interviews: An Introduction to Qualitative Research Interviewing. Thousand Oaks, CA: SAGE.

Kvale, Steinar. 2002. "The Social Construction of Validity." Pp. 299–325 in The Qualitative Inquiry Reader, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Labaw, Patricia J. 1980. Advanced Questionnaire Design. Cambridge, MA: ABT Books.

Lacey, John H., Tara Kelley-Baker, Robert B. Voas, Eduardo Romano, C. Debra Furr-Holden, Pedro Torres, and Amy Berning. 2011. "Alcohol- and Drug-Involved Driving in the United States: Methodology for the 2007 National Roadside Survey." Evaluation Review 35:319–353.

Lange, Matthew. 2013. Comparative–Historical Methods. Thousand Oaks, CA: SAGE.

Langford, Terri. 2000. "Census Workers in Dallas Find the Well-Off Hard to Count." The Boston Globe, June 1, p. A24.

LaRossa, Ralph. 1995. "Parenthood in Early Twentieth- Century America Project (PETCAP), 1900–1944" [Computer file]. Atlanta, GA: Georgia State University [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1997.

Larson, Calvin J. 1993. Pure and Applied Sociological Theory: Problems and Issues. New York: Harcourt Brace Jovanovich.

Lathrop, Barnes F. 1968. "History From the Census Returns." Pp. 79–101 in Sociology and History: Methods, edited by Seymour Martin Lipset and Richard Hofstadter. New York: Basic Books.

Latour, Francie. 2002. "Marching Orders: After 10 Years, State Closes Prison Boot Camp." Boston Sunday Globe, June 16, pp. B1, B7.

Lavin, Michael R. 1994. Understanding the 1990 Census: A Guide for Marketers, Planners, Grant Writers and Other Data Users. Kenmore, NY: Epoch Books.

Lavrakas, Paul J. 1987. Telephone Survey Methods: Sampling, Selection, and Supervision. Newbury Park, CA: SAGE.

Lavrakas, Paul J., Grant Benson, Stephen Blumberg, Trent Buskirk, Ismael Flores Cervantes, Leah Christian et al. 2017. Report from the AAPOR Task Force on "The Future of U.S. General Population Telephone Survey Research." Retrieved http://www.aapor.org/getattachment/Education-Resources/Reports/Future-of-Telephone-Survey-Research-Report.pdf.aspx

Layte, Richard and Christopher T. Whelan. 2003. "Moving In and Out of Poverty: The Impact of Welfare Regimes on Poverty Dynamics in the EU." European Societies 5:167–191.

Leakey, Tricia, Kevin B. Lunde, Karin Koga, and Karen Glanz. 2004. "Written Parental Consent and the Use of Incentives in a Youth Smoking Prevention Trial: A Case Study From Project SPLASH." American Journal of Evaluation 25:509–523.

LeDuc, Lawrence, Richard G. Niemi, and Pippa Norris (Eds.). 1996. Comparing

Democracies: Elections and Voting in Global Perspective. Thousand Oaks, CA: SAGE.

Lehrer, Jonah. 2010. “The Truth Wears Off: Is There Something Wrong With the Scientific Method?” The New Yorker, December 13:52–57.

Lempert, Richard. 1989. “Humility Is a Virtue: On the Publicization of Policy-Relevant Research.” Law & Society Review 23:146–161.

Lempert, Richard and Joseph Sanders. 1986. An Invitation to Law and Social Science: Desert, Disputes, and Distribution. New York: Longman.

Levine, James P. 1976. “The Potential for Crime Overreporting in Criminal Victimization Surveys.” Criminology 14:307–330.

Levinson, Martin P. 2010. “Accountability to Research Participants: Unresolved Dilemmas and Unravelling Ethics.” Ethnography and Education 5:193–207.

Levitas, Ruth and Will Guy. 1996. “Introduction.” Pp. 1–6 in Interpreting Official Statistics, edited by Ruth Levitas and Will Guy. New York: Routledge.

Levitt, Heidi M., Rebecca Todd Swanger, and Jenny B. Butler. 2008. “Male Perpetrators’ Perspectives on Intimate Partner Violence, Religion, and Masculinity.” Sex Roles 58:435–448.

Levy, Paul S. and Stanley Lemeshow. 1999. Sampling of Populations: Methods and Applications, 3rd ed. New York: Wiley.

Lewin, Tamar. 2001a. “Income Education Is Found to Lower Risk of New Arrest.” The New York Times, November 16, p. A18.

Lewin, Tamar. 2001b. “Surprising Result in Welfare-to-Work Studies.” The New York Times, July 31, p. A16.

Lewis, Kevin. 2015. “Three Fallacies of Digital Footprints.” Big Data & Society (July-December):1–4.

Lewis, Kevin, Nicholas Christakis, Marco Gonzalez, Jason Kaufman, and Andreas Wimmer. 2008. “Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com.” Social Networks 30(4):330–342.

Lewis-Beck, Michael S., Alan Bryman, and Tim Futing Liao (Eds.). 2004. The SAGE Encyclopedia of Social Science Research Methods, vol. 1. Thousand Oaks, CA: SAGE.

Li, Jun. 2008. “Ethical Challenges in Participant Observation: A Reflection on Ethnographic Fieldwork.” The Qualitative Report 13:100–115.

Liberatore, Stacy. 2016. “What Happens in an Internet Second: 54,907 Google Searches, 7,252 Tweets, 125,406 YouTube Video Views and 2,501,018 Emails Sent.” Daily Mail (June 27). Retrieved from http://www.dailymail.co.uk/sciencetech/article-3662925/What-happens-internet-second-54-907-Google-searches-7-252-tweets-125-406-YouTube-video-views-2-501-018-emails-sent.html

Lieberson, Stanley. 1985. Making It Count: The Improvement of Social Research and Theory. Berkeley: University of California Press.

Lieberson, Stanley. 1991. “Small N’s and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases.” Social Forces 70:307–320.

Lillard, Lee A. and Constantijn W. A. Panis. 1998. “Panel Attrition From the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality.” The Journal of Human Resources 33:437–457.

Lincoln, Yvonna S. 2009. “Ethical Practices in Qualitative Research.” Pp. 150–169 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Lindsay, Sally, Simon Smith, Frances Bell, and Paul Bellaby. 2007. “Tackling the Digital Divide: Exploring the Impact of ICT on Managing Heart Conditions in a Deprived Area.” Information, Communication & Society 10:95–114.

Ling, Rich and Gitte Stald. 2010. “Mobile Communities: Are We Talking About a Village, a Clan, or a Small Group?” American Behavioral Scientist 53(8):113–1147.

Link, Bruce G., Jo C. Phelan, Ann Stueve, Robert E. Moore, Michaeline Brenahan, and Elmer L. Struening. 1996. “Public Attitudes and Beliefs About Homeless People.” Pp. 143–148 in Homelessness in America, edited by Jim Baumohl. Phoenix, AZ: Oryx.

Lipset, Seymour Martin. 1968. Revolution and Counterrevolution. New York: Basic Books.

Lipset, Seymour Martin, Martin Trow, and James Coleman. 1956. Union Democracy: The Internal Politics of the International Typographical Union. New York: Doubleday Anchor.

Lipsey, Mark W. and David B. Wilson. 2001. Practical Meta-Analysis. Thousand Oaks, CA: SAGE.

Lipsky, Michael. 1980. Street-Level Bureaucracy. New York: Russell Sage Foundation.

Liptak, Kevin. 2012. “Report Shows Turnout Lower Than 2008 and 2004.” CNN. Retrieved from http://politicalticker.blogs.cnn.com/2012/11/08/report-shows-turnout-lower-than-2008-and-2004

Litwin, Mark S. 1995. How to Measure Survey Reliability and Validity. Thousand Oaks, CA: SAGE.

Loader, Brian D., Steve Muncer, Roger Burrows, Nicolas Pleace, and Sarah Nettleton. 2002. “Medicine on the Line? Computer-Mediated Social Support and Advice for People With Diabetes.” International Journal of Social Welfare 11:53–65.

Locke, Lawrence F., Stephen J. Silverman, and Waneen Wyrick Spirduso. 1998. Reading and Understanding Research. Thousand Oaks, CA: SAGE.

Locke, Lawrence F., Waneen Wyrick Spirduso, and Stephen J. Silverman. 2000. Proposals That Work: A Guide for Planning Dissertations and Grant Proposals, 4th ed. Thousand Oaks, CA: SAGE.

Lockwood, Daniel. 1996. Violent Incidents Among Selected Public School Students in Two Large Cities of the South and the Southern Midwest, 1995: [United States] [Computer file]. ICPSR version. Atlanta, GA: Clark Atlantic University [producer], 1996. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1998.

Lofland, John and Lyn H. Lofland. 1984. Analyzing Social Settings: A Guide to Qualitative Observation and Analysis, 2nd ed. Belmont, CA: Wadsworth.

Lund, Laura and William E. Wright. 1994. "*Mitofsky–Waksberg vs.* Screened Random Digit Dial: Report on a Comparison of the Sample Characteristics of Two RDD Survey Designs." Presented at the Center for Disease Control's 11th Annual Behavioral Risk Factor Survey Conference, Atlanta, GA, June.

Luxardo, Natalia, Graciela Colombo, and Gabriela Iglesias. 2011. "Methodological and Ethical Dilemmas Encountered During Field Research of Family Violence Experienced by Adolescent Women in Buenos Aires." The Qualitative Report 16: 984–1000.

Lyall, Sarah. 2004. "Does Queen Get Her Mail on Time? You've Got to Wonder." The New York Times, May 28, p. A4.

Lyberg, Lars, and Patricia Dean. 1992. "Methods for Reducing Nonresponse Rates: A Review." Annual Meeting of the American Association for Public Opinion Research. St. Petersburg, Florida.

Lynch, Michael and David Bogen. 1997. "Sociology's Asociological 'Core': An Examination of Textbook Sociology in Light of the Sociology of Scientific Knowledge." American Sociological Review 62:481–493.

Mabry, Linda. 2008. “Case Study in Social Research.” Pp. 214–227 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

MacDougall, Colin and Frances Baum. 1997. “The Devil’s Advocate: A Strategy to Avoid Groupthink and Stimulate Discussion in Focus Groups.” Qualitative Health Research 7:532–541.

Madden, Raymond. 2010. Being Ethnographic: A Guide to the Theory and Practice of Ethnography. Thousand Oaks, CA: SAGE.

Mahoney, James. 2001. The Legacies of Liberalism: Path Dependence and Political Regimes in Central America. New York: Cambridge University Press.

Majchrzak, Ann and M. Lynne Markus. 2014. Methods for Policy Research: Taking Socially Responsible Action, 2nd ed. Thousand Oaks, CA: SAGE.

Mangione, Thomas W. 1995. Mail Surveys: Improving the Quality. Thousand Oaks, CA: SAGE.

Manza, Jeff, Clem Brooks, and Michael Sauder. 2005. “Money, Participation, and Votes: Social Cleavages and Electoral Politics.” Pp. 201–226 in The Handbook of Political Sociology: States, Civil Societies, and Globalization, edited by Thomas Janoski, Robert R. Alford, Alexander M. Hicks, and Mildred A. Schwartz. New York: Cambridge University Press.

Margolis, Eric. 2004. “Looking at Discipline, Looking at Labour: Photographic Representations of Indian Boarding Schools.” Visual Studies 19:72–96.

Marini, Margaret Mooney and Burton Singer. 1988. “Causality in the Social Sciences.” Pp. 347–409 in Sociological Methodology, vol. 18, edited by Clifford C. Clogg. Washington, DC: American Sociological Association.

Mark, Melvin M. and Chris Gamble. 2009. “Experiments, Quasi-Experiments, and

Ethics." Pp. 198–213 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Markham, Annette N. 2008. "The Methods, Politics, and Ethics of Representation in Online Ethnography." Pp. 247–284 in Collecting and Interpreting Qualitative Materials, 3rd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Markoff, John. 2005. "Transitions to Democracy." Pp. 384–403 in The Handbook of Political Sociology: States, Civil Societies, and Globalization, edited by Thomas Janoski, Robert R. Alford, Alexander M. Hicks, and Mildred A. Schwartz. New York: Cambridge University Press.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." American Sociological Review 52:122–131.

Marshall, Gary D. and Philip G. Zimbardo. 1979. "Affective Consequences of Inadequately Explained Physiological Arousal." Journal of Personality and Social Psychology 37:970–988.

Martin, Lawrence L. and Peter M. Kettner. 1996. Measuring the Performance of Human Service Programs. Thousand Oaks, CA: SAGE.

Martin, Linda G. and Kevin Kinsella. 1995. "Research on the Demography of Aging in Developing Countries." Pp. 356–403 in Demography of Aging, edited by Linda G. Martin and Samuel H. Preston. Washington, DC: National Academy Press.

Marx, Karl. 1967. Capital: A Critique of Political Economy. New York: International Publishers.

Marx, Karl and Friedrich Engels. 1961. "The Communist Manifesto." Pp. 13–44 in Essential Works of Marxism, edited by Arthur P. Mendel. New York: Bantam.

Matejka, Justin, Michael Glueck, Tovi Grossman and George Fitzmaurice. 2016. "The

Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales.” CHI. http://www.dgp.toronto.edu/~tovi/papers/vas.pdf

Matt, Georg E. and Thomas D. Cook. 1994. “Threats to the Validity of Research Syntheses.” Pp. 503–520 in The Handbook of Research Synthesis, edited by Harris Cooper and Larry V. Hedges. New York: Russell Sage Foundation.

Maxwell, Joseph A. 2005. Qualitative Research Design: An Interactive Approach, 2nd ed. Thousand Oaks, CA: SAGE.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt.

Maynard, Douglas W., Jeremy Freese, and Nora Cate Schaeffer. 2010. “Calling for Participation: Requests, Blocking Moves, and Rational (Inter)action in Survey Introductions.” American Sociological Review 75:791–814.

Mayrl, Damon, Ben Moodie, Jon Norman, Jodi Short, Sarah Staveteig, and Cinzia Solari. 2004. “A Theory of Relativity.” Contexts 3:10.

McCarter, Susan A. 2009. “Legal and Extralegal Factors Affecting Minority Overrepresentation in Virginia’s Juvenile Justice System: A Mixed-Method Study.” Child and Adolescent Social Work Journal 26:533–544.

McGeeney, Kyley and Scott Keeter. 2014. “Pew Research Increases Share of Interviews Conducted by Cellphone.” Pew Research Center, January 15. Retrieved from www.pewresearch.org/fact-tank/2014/01/15/pew-research-increases-share-of-interviews-conducted-by-cellphone

McIntyre, Alice. 2008. Participatory Action Research. Thousand Oaks, CA: SAGE.

McLellan, A. Thomas, Lester Luborsky, John Cacciola, Jeffrey Griffith, Frederick Evans, Harriet L. Barr, and Charles P. O’Brien. 1985. “New Data From the Addiction Severity Index: Reliability and Validity in Three Centers.” The Journal of Nervous and Mental

Disease 173(7):412–423.

McPhail, Clark and John McCarthy. 2004. “Who Counts and How: Estimating the Size of Protests.” Contexts 3:12–18.

McPherson, Miller, Lynn Smith-Lovin, and Matthew E. Brashears. 2006. “Social Isolation in America: Changes in Core Discussion Networks Over Two Decades.” American Sociological Review 71:353–375.

Medway, Rebecca L. and Jenna Fulton. 2012. “When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates.” Public Opinion Quarterly 76:733–746.

Meghani, S. H., Brooks, J., Gipson-Jones, T., Waite, R., Whitefield-Harris, L., Deatrick, J. 2009. “Patient–Provider Race-Concordance: Does it Matter in Improving Minority Patients’ Health Outcomes?” Ethnicity & Health 14(1):107–130.

Menikoff, Jerry, Julie Kaneshiro, and Ivor Pritchard. 2017. “The Common Rule, Updated.” The New England Journal of Medicine 376(7):613–615.

Mertens, Donna M. 2012. “Transformative Mixed Methods: Addressing Inequities.” American Behavioral Scientist 56:802–813.

Merton, Robert K., Marjorie Fiske, and Patricia L. Kendall. 1956. The Focused Interview. Glencoe, IL: Free Press.

Michel, Franck. 2012. “How Many Photos Are Uploaded to Flickr Every Day, Month, Year.” Flickr, March 20. Retrieved from https://www.flickr.com/photos/franckmichel/6855169886

Mieczkowski, Tom. 1997. “Hair Assays and Urinalysis Results for Juvenile Drug Offenders.” National Institute of Justice Research Preview. Washington, DC: U.S. Department of Justice.

Milbrath, Lester and M. L. Goel. 1977. Political Participation, 2nd ed. Chicago: Rand McNally.

Miles, Matthew B. and A. Michael Huberman. 1994. Qualitative Data Analysis, 2nd ed. Thousand Oaks, CA: SAGE.

Miles, Matthew B., A. Michael Huberman, and Johnny Soldaña. 2014. Qualitative Data Analysis: A Methods Sourcebook, 3rd ed. Thousand Oaks: SAGE.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." Journal of Abnormal and Social Psychology 67:371–478.

Milgram, Stanley. 1964. "Issues in the Study of Obedience: A Reply to Baumrind." American Psychologist 19:848–852.

Milgram, Stanley. 1965. "Some Conditions of Obedience and Disobedience to Authority." Human Relations 18:57–76.

Milgram, Stanley. 1974. Obedience to Authority: An Experimental View. New York: Harper & Row.

Milgram, Stanley. 1977. "Subject Reaction: The Neglected Factor in the Ethics of Experimentation." Hastings Law Review, October, pp. 19–23 (as cited in Cave, Emm and Soren Holm. 2003. "Milgram and Tuskegee—Paradigm Research Projects in Bioethics." Health Care Analysis 11:27–40).

Milgram, Stanley. 1992. The Individual in a Social World, 2nd ed. New York, NY: McGraw-Hill.

Mill, John Stuart. 1872. A System of Logic: Ratiocinative and Inductive, 8th ed., vol. 2. London: Longmans, Green, Reader, & Dyer.

Miller, Arthur G. 1986. The Obedience Experiments: A Case Study of Controversy in Social Science. New York: Praeger.

Miller, Delbert C. and Neil J. Salkind. 2002. Handbook of Research Design and Social Measurement, 6th ed. Thousand Oaks, CA: SAGE.

Miller, JoAnn. 2003. “An Arresting Experiment: Domestic Violence Victim Experiences and Perceptions.” Journal of Interpersonal Violence 18:695–716.

Miller, Susan. 1999. Gender and Community Policing: Walking the Talk. Boston: Northeastern University Press.

Miller, Ted R. and Delia Hendrie. 2008. Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis, DHHS Pub. No. (SMA) 07-4298. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.

Miller, Walter B. 1992. Crime by Youth Gangs and Groups in the United States. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Miller, William L. and Benjamin F. Crabtree. 1999a. “Clinical Research: A Multimethod Typology and Qualitative Roadmap.” Pp. 3–30 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Miller, William L. and Benjamin F. Crabtree. 1999b. “The Dance of Interpretation.” Pp. 127–143 in Doing Qualitative Research, edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Miller, William L. and Benjamin F. Crabtree. 1999c. “Depth Interviewing.” Pp. 89–107 in Doing Qualitative Research, 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: SAGE.

Mills, C. Wright. 1959. The Sociological Imagination. New York: Oxford University Press.

Minkler, Meredith. 2000. “Using Participatory Action Research to Build Healthy Communities.” Public Health Reports 115:191–197.

Mirowsky, John. 1999. Aging, Status, and the Sense of Control: Competing Continuation (Proposal to the National Institute of Aging). Urbana: University of Illinois Press.

Mirowsky, John and Catherine E. Ross. 1999. “Economic Hardship Across the Life Course.” American Sociological Review 64:548–569.

Mirowsky, John and Catherine E. Ross. 2001. Aging, Status, and the Sense of Control (ASOC), 1995, 1998, 2001 [United States] Questionnaire (ICPSR 3334). Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Mirowsky, John and Catherine E. Ross. 2003. Education, Social Status, and Health. New York: Aldine de Gruyter.

Mitchell, Richard G. Jr. 1993. Secrecy and Fieldwork. Newbury Park, CA: SAGE.

Moe, Angela M. 2007. “Silenced Voices and Structural Survival—Battered Women’s Help Seeking.” Violence Against Women 13(7):676–699.

Moe, Hallvard and Anders Olof Larsson. 2012. “Methodological and Ethical Challenges Associated With Large-Scale Analyses of Online Political Communication.” NORDICOM Review: Nordic Research on Media and Communication 33(1):117–124.

Mohr, Lawrence B. 1992. Impact Analysis for Program Evaluation. Newbury Park, CA: SAGE.

Monkkonen, Eric H. 1994. “Introduction.” Pp. 1–8 in Engaging the Past: The Uses of History Across the Social Sciences. Durham, NC: Duke University Press.

Mooney, Christopher Z. and Mei Hsien Lee. 1995. “Legislating Morality in the American States: The Case of Abortion Regulation Reform.” American Journal of Political Science 39:599–627.

Moore, Barrington Jr. 1966. Social Origins of Democracy and Dictatorship: Lord and

Peasant in the Making of the Modern World. Boston: Beacon Press.

Moore, Spencer, Mark Daniel, Laura Linnan, Marci Campbell, Salli Benedict, and Andrea Meier. 2004. “After Hurricane Floyd Passed: Investigating the Social Determinants of Disaster Preparedness and Recovery.” Family and Community Health 27: 204–217.

Morgan, David L. 2014. Integrating Qualitative & Quantitative Methods: A Pragmatic Approach. Thousand Oaks, CA: SAGE.

Morrill, Calvin, Christine Yalda, Madeleine Adelman, Michael Musheno, and Cindy Bejarano. 2000. “Telling Tales in School: Youth Culture and Conflict Narratives.” Law & Society Review 34:521–565.

Mosher, Clayton J., Terance D. Miethe, and Dretha M. Phillips. 2002. The Mismeasure of Crime. Thousand Oaks, CA: SAGE.

Muhr, Thomas and Susanne Friese. 2004. User’s Manual for ATLAS.ti 5.0, 2nd ed. Berlin: Scientific Software Development.

Myers, Steven Lee. 2002. “Russia Takes Stock of a Nation’s Transformation.” The New York Times, September 29, p. 3.

Myers, Steven Lee. 2010. “Delays in a Head Count Keep Crucial Numbers a Matter of Guesswork.” The New York Times, December 7, p. A10.

Nagourney, Adam. 2002. “Cellphones and Caller ID Are Making It Harder for Pollsters to Pick a Winner.” The New York Times, November 5, p. A20.

Nakonezny, Paul A., Rebecca Reddick, and Joseph Lee Rodgers. 2004. “Did Divorces Decline After the Oklahoma City Bombing?” Journal of Marriage and Family 66:90–100.

Narayan, Sowmya and Jon A. Krosnick. 1996. “Education Moderates Some Response Effects in Attitude Measurement.” Public Opinion Quarterly 60:58–88.

National Center for Bioethics in Research and Health Care. n.d. About the USPHS Syphilis Study. Retrieved from http://tuskegeebioethics.org/about-the-usphs-syphilis-study/

National Geographic Society. 2000. Survey 2000. Retrieved from http://survey2000.nationalgeographic.com

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 1994. “Alcohol-Related Impairment.” Alcohol Alert 25 (July):1–5.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 1997. “Alcohol Metabolism.” Alcohol Alert 35(January):1–4.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). n.d. “Alcohol Use Disorder.” Retrieved from https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 2013. NIAAA Recognizes Alcohol Awareness Month. Retrieved from www.niaaa.nih.gov/news-events/alcohol-awareness-month-2013

National Oceanic and Atmospheric Administration (NOAA). 2005. Hurricane Katrina. Washington, DC: U.S. Department of Commerce. Retrieved from www.ncdc.noaa.gov/extremeevents/specialreports/Hurricane-Katrina.pdf

National Oceanic and Atmospheric Administration (NOAA). 2017. Hurricane Harvey. Retrieved from http://www.nhc.noaa.gov/archive/2017/HARVEY.shtml

National Opinion Research Center (NORC). 2011. General Social Survey (GSS). Chicago: NORC at the University of Chicago. Retrieved from www.norc.org/Research/Projects/Pages/general-social-survey.aspx

National Opinion Research Center (NORC). 2014. General Social Survey. Retrieved from www3.norc.org/GSS+Website/Publications/GSS+Questionnaires

National Opinion Research Center (NORC). 2016. Get the Data. Retrieved from http://gss.norc.org/Get-The-Data

Needleman, Carolyn. 1981. “Discrepant Assumptions in Empirical Research: The Case of Juvenile Court Screening.” Social Problems 28 (February):247–262.

Neuendorf, Kimberly A. 2002. The Content Analysis Guidebook. Thousand Oaks, CA: SAGE.

Newbury, Darren. 2005. “Editorial: The Challenge of Visual Studies.” Visual Studies 20:1–3.

Newport, Frank. 1992. “Look at Polls as a Fever Chart of the Electorate.” Letter to the Editor, The New York Times, November 6, p. A28.

Nie, Norman H. and Lutz Erbring. 2002. Internet and Society: A Preliminary Report. Palo Alto, CA: Stanford Institute for the Quantitative Study of Society.

Nordanger, Dag. 2007. “Discourses of Loss and Bereavement in Tigray, Ethiopia.” Culture, Medicine and Psychiatry 31:173–194.

Norris, Pippa. 2004. “The Bridging and Bonding Role of Online Communities.” Pp. 31–41 in Society Online: The Internet in Context, edited by Philip N. Howard and Steve Jones. Thousand Oaks, CA: SAGE.

Nunberg, Geoffrey. 2002. “The Shifting Lexicon of Race.” The New York Times, December 22, p. WK3.

O’Brien, Daniel Tumminelli and Christopher Winship. 2017. “The Gains of Greater Granularity: The Presence and Persistence of Problem Properties in Urban Neighborhoods.” Journal of Quantitative Criminology 33:649–674.

Ó Dochartaigh, Niall. 2002. The Internet Research Handbook: A Practical Guide for Students and Researchers in the Social Sciences. Thousand Oaks, CA: SAGE.

Office of Inspector General, U.S. Postal Service. 2017. Delivery and Customer Service Operations–Bronx, NY. Retrieved from https://www.uspsoig.gov/sites/default/files/document-library-files/2017/DR-AR-17-004.pdf

Ogburn, William F. 1930. "The Folkways of a Scientific Sociology," Scientific Monthly 30:300–306.

Olson, Kristen, Jolene D. Smyth, and Heather M. Wood. 2012. "Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination." Public Opinion Quarterly 76:611–635.

Ondersma, Steven J., Joanne Martin, Beverly Fortson, Daniel J. Whitaker, Shannon Self-Brown, Jessica Beatty, Amy Loree, David Bard, and Mark Chaffin. 2017. "Technology to Augment Early Home Visitation for Child Maltreatment Prevention: A Pragmatic Randomized Trial." Child Maltreatment. doi:10.1177/1077559517729890

Onoye, Jane M., Deborah A. Goebert, and Stephanie T. Nishimura. 2012. "Use of Incentives and Web-Based Administration for Surveying Student Alcohol and Substance Use in an Ethnically Diverse Sample." Journal of Substance Use 17(1):61–71.

Orcutt, James D. and J. Blake Turner. 1993. "Shocking Numbers and Graphic Accounts: Quantified Images of Drug Problems in the Print Media." Social Problems 49(May):190–206.

Organisation for Economic Co-operation and Development. 2017. "SF3.1: Marriage and Divorce Rates." Retrieved from https://www.oecd.org/els/family/SF_3_1_Marriage_and_divorce_rates.pdf

Orr, Larry L. 1999. Social Experiments: Evaluating Public Programs With Experimental Methods. Thousand Oaks, CA: Sage.

Orshansky, Mollie. 1977. "Memorandum for Daniel P. Moynihan. Subject: History of the Poverty Line." Pp. 232–237 in The Measure of Poverty. Technical Paper I: Documentation of Background Information and Rationale for Current Poverty Matrix, edited by Mollie Orshansky. Washington, DC: U.S. Department of Health, Education, and Welfare.

Orwin, Robert G. 1994. “Evaluating Coding Decisions.” Pp. 138–162 in The Handbook of Research Synthesis, edited by Harris Cooper and Larry V. Hodges. New York: Russell Sage Foundation.

Ousey, Graham C. and Matthew R. Lee. 2004. “Investigating the Connections Between Race, Illicit Drug Markets, and Legal Violence, 1984–1997.” Journal of Research in Crime and Delinquency 41:352–383.

Pagnini, Deanna L. and S. Philip Morgan. 1996. “Racial Differences in Marriage and Childbearing: Oral History Evidence From the South in the Early Twentieth Century.” American Journal of Sociology 101:1694–1715.

Paige, Jeffery M. 1999. “Conjuncture, Comparison, and Conditional Theory in Macrosocial Inquiry.” American Journal of Sociology 105:781–800.

Paika, Anthony and Kenneth Sanchagrina. 2013. “Social Isolation in America: An Artifact.” American Sociological Review 78(3):339–360.

Pan, Yuling and Manuel de la Puente. 2005. Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on How the Guideline Was Developed (Research Report Series, Survey Methodology #2005-06). Washington, DC: Statistical Research Division, U.S. Census Bureau.

Panagopoulos, Costas. 2008. “Poll Accuracy in the 2008 Presidential Election.” Retrieved from www.fordham.edu/images/academics/graduate_schools/gsas/elections_and_campaign_/p

Papineau, David. 1978. For Science in the Social Sciences. London: Macmillan.

Parker, Kim and Renee Stepler. 2017. “As U.S. Marriage Rate Hovers at 50%, Education Gap in Marital Status Widens.” Pew Research Center. Retrieved from http://www.pewresearch.org/fact-tank/2017/09/14/as-u-s-marriage-rate-hovers-at-50-education-gap-in-marital-status-widens/

Parks, Kathleen A., Ann M. Pardi, and Clara M. Bradizza. 2006. “Collecting Data on Alcohol Use and Alcohol-Related Victimization: A Comparison of Telephone and Web-Based Survey Methods.” Journal of Studies on Alcohol 67:318–323.

Parlett, Malcolm and David Hamilton. 1976. “Evaluation as Illumination: A New Approach to the Study of Innovative Programmes.” Pp. 140–157 in Evaluation Studies Review Annual, vol. 1, edited by G. Glass. Beverly Hills, CA: SAGE.

Pate, Antony M. and Edwin E. Hamilton. 1992. “Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment.” American Sociological Review 57(October):691–697.

Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. “Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault.” Law & Society Review 31(1):163–204.

Patrick, Megan E. and Yvonne M. Terry-McElrath. 2017. “High Intensity Drinking by Underage Young Adults in the United States.” Addiction 112:82–93.

Patterson, Orlando. 1997. “The Race Trap.” The New York Times, July 11, p. A25.

Patton, Michael Quinn. 2002. Qualitative Research & Evaluation Methods, 3rd ed. Thousand Oaks, CA: SAGE.

Patulny, Roger and Claire Seaman. 2017. “‘I’ll Just Text You’: Is Face-to-Face Social Contact Declining in a Mediated World?” Journal of Sociology 53(2):285–302.

Pauwels, Luc. 2010. “Visual Sociology Reframed: An Analytical Synthesis and Discussion of Visual Methods in Social and Cultural Research.” Sociological Methods & Research 38:545–581.

Paxton, Pamela. 2002. “Social Capital and Democracy: An Interdependent Relationship.” American Sociological Review 67:254–277.

Pearson, Michael, Helen Sweeting, Patrick West, Robert Young, Jacki Gordon, and Katrina Turner. 2006. “Adolescent Substance Use in Different Social and Peer Contexts: A Social Network Analysis.” Drugs: Education, Prevention and Policy 13:519–536.

Pedulla, David S. 2016. “Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories.” American Sociological Review 81(2):262–289.

Pepinsky, Harold E. 1980. “A Sociologist on Police Patrol.” Pp. 223–234 in Fieldwork Experience: Qualitative Approaches to Social Research, edited by William B. Shaffir, Robert A. Stebbins, and Allan Turowetz. New York: St. Martin’s Press.

Percac-Lima, Sanja, Leslie S. Aldrich, Gloria B. Gamba, Adriana M. Bearse, and Steven J. Atlas. 2010. “Barriers to Follow-Up of an Abnormal Pap Smear in Latina Women Referred for Colposcopy.” Journal of General Internal Medicine 25(11):1198–1204.

Perrin, Andrew W. and Maeve Duggan. 2015. “Americans’ Internet Access: 2000–2015.” Pew Research Center, Internet & Technology, June 26. Retrieved from http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/

Perry, Gina. 2013. Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments. New York: New Press.

Peterson, Robert A. 2000. Constructing Effective Questionnaires. Thousand Oaks, CA: SAGE.

Petrosino, Anthony, Robert F. Boruch, Haluk Soydan, Lorna Duggan, and Julio Sanchez-Meca. 2001. “Meeting the Challenges of Evidence-Based Policy: The Campbell Collaboration.” Annals of the American Academy of Political and Social Science 578(14):14–17.

Pew Hispanic Center. 2008. Statistical Portrait of the Foreign-Born Population in the United States, 2006. Statistical tables from 2000 Census and 2006 American Community Survey, Retrieved from http://pewhispanic.org/factsheets/factsheet.php?FactsheetID=36

Pew Research Center. 2014. Political Polarization in the American Public. Washington, DC: Pew Research Center.

Pew Research Center. 2017a. The American Trends Panel Survey Methodology. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/09/14122154/W28-American-Trends-Panel-Survey-Methodology_CHECKED.pdf

Pew Research Center. 2017b. 2017 Pew Research Center's American Trends Panel. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/09/14122053/W28-Topline_marriage_blog-post_FINAL.pdf.

Phillips, David P. 1982. "The Impact of Fictional Television Stories on U.S. Adult Fatalities: New Evidence on the Effect of the Mass Media on Violence." American Journal of Sociology 87:1340–1359.

Piliavin, Jane Allyn and Irving M. Piliavin. 1972. "Effect of Blood on Reactions to a Victim." Journal of Personality and Social Psychology 23:353–361.

Pink, Sarah, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi. 2016. Digital Ethnography: Principles and Practice. Thousand Oaks, CA: SAGE.

Plano Clark, Vicki L. and Nataliya V. Ivankova. 2016. Mixed Methods Research: A Guide to the Field. Thousand Oaks, CA: SAGE.

Plessy v. Ferguson, 163 U.S. 537 (1896).

Pollner, Melvin and Richard E. Adams. 1994. "The Interpersonal Context of Mental

Health Interviews." Journal of Health and Social Behavior 35:283–290.

Popper, Sir Karl. 1968/1934. The Logic of Scientific Discovery. New York: Harper & Row.

Porter, Stephen R. and Michael E. Whitcomb. 2003. "The Impact of Contact Type on Web Survey Response Rates." Public Opinion Quarterly 67:579–588.

Posavac, Emil J. and Raymond G. Carey. 1997. Program Evaluation: Methods and Case Studies, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Presley, Cheryl A., Philip W. Meilman, and Rob Lyerla. 1994. "Development of the Core Alcohol and Drug Survey: Initial Findings and Future Directions." Journal of American College Health 42:248–255.

Presser, Stanley and Johnny Blair. 1994. "Survey Pretesting: Do Different Methods Produce Different Results?" Sociological Methodology 24:73–104.

Presser, Stanley, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. 2004. "Methods for Testing and Evaluating Survey Questions." Public Opinion Quarterly 68:109–130.

Price, Richard H., Michelle Van Ryn, and Amiram D. Vinokur. 1992. "Impact of a Preventive Job Search Intervention on the Likelihood of Depression Among the Unemployed." Journal of Health and Social Behavior 33 (June):158–167.

Puchta, Claudia and Jonathan Potter. 2004. Focus Group Practice. Thousand Oaks, CA: Sage Publications.

Punch, Maurice. 1994. "Politics and Ethics in Qualitative Research." Pp. 83–97 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Putnam, Israel. 1977. "Poverty Thresholds: Their History and Future Development." Pp.

272–283 in The Measure of Poverty. Technical Paper I: Documentation of Background Information and Rationale for Current Poverty Matrix, edited by Mollie Orshansky. Washington, DC: U.S. Department of Health, Education, and Welfare.

Pyrczak, Fred. 2005. Evaluating Research in Academic Journals: A Practical Guide to Realistic Evaluation, 3rd ed. Glendale, CA: Pyrczak.

Quealy, Kevin. 2017. "The Cost of Hurricane Harvey: Only One Recent Storm Comes Close." The New York Times, September 1.

Radloff, Lenore. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." Applied Psychological Measurement 1:385–401.

Ragin, Charles C. 1987. The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. Berkeley: University of California Press.

Ragin, Charles C. 1994. Constructing Social Research. Thousand Oaks, CA: SAGE.

Ragin, Charles C. 2000. Fuzzy-Set Social Science. Chicago: University of Chicago Press.

Rainie, Lee, Joanna Brenner, and Kristen Purcell. 2012. Photos and Videos as Social Currency Online. Washington, DC: Pew Internet & American Life Project. Retrieved from http://pewinternet.org/Reports/2012/Online-Pictures.aspx

Rainie, Lee and John Horrigan. 2005. A Decade of Adoption: How the Internet Has Woven Itself Into American Life. Pew Internet & American Life Project. Retrieved from the Pew Internet & American Life Project, www.pewinternet.org/PPF/r/148/report_display.asp (PDF version).

Rainie, Lee, Aaron Smith, and Maeve Duggan. 2013. Coming and Going on Facebook. Washington, DC: Pew Internet & American Life Project.

Randall, Ann. 2012. "Beneficial Interview Effects in Virtual Worlds: A Case Study." Pp. 131–149 in Cases in Online Interview Research, edited by Janet Salmons. Thousand

Oaks, CA: SAGE.

Rank, Steven G. and Cardell K. Jacobson. 1977. “Hospital Nurses’ Compliance With Medication Overdose Orders: A Failure to Replicate.” Journal of Health and Social Behavior 18:188–193.

Rashbaum, William K. 2002. “Reasons for Crime Drop in New York Elude Many.” The New York Times, November 29, p. A28.

Raudenbush, Stephen W. and Robert J. Sampson. 1999. “Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods.” Sociological Methodology 29:1–41.

Ravitch, Sharon M. and Matthew Riggan. 2016. Reason & Rigor: How Conceptual Frameworks Guide Research, 2nd ed. Thousand Oaks, CA: SAGE.

Regan, Michael D. 2016. “What Does Voter Turnout Tell Us About the 2016 Election?” PBS Newshour, November 20. Retrieved from http://www.pbs.org/newshour/updates/voter-turnout-2016-elections/

Reips, Ulf-Dietrich and Christoph Neuhaus. 2002. “WEXTOR: A Web-Based Tool for Generating and Visualising Experimental Designs and Procedures.” Behavior Research Methods Instruments & Computers 36(2):304–311.

Reisman, David. 1950/1969. The Lonely Crowd: A Study of the Changing American Character. New Haven, CT: Yale University Press.

Reiss, Albert J. Jr. 1971a. The Police and the Public. New Haven, CT: Yale University Press.

Reiss, Albert J. Jr. 1971b. “Systematic Observations of Natural Social Phenomena.” Pp. 3–33 in Sociological Methodology, vol. 3, edited by Herbert Costner. San Francisco: Jossey-Bass.

Rele, J. R. 1993. "Demographic Rates: Birth, Death, Marital, and Migration." Pp. 2-1–2-26 in Readings in Population Research Methodology. Vol. 1, Basic Tools, edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United National Population Fund.

Reuter, Peter, Rosalie Liccardo Pacula, and Jonathan P. Caulkins. 2010. "Addiction Research Centres and the Nurturing of Creativity: RAND's Drug Policy Research Center." Addiction 106:253–259.

Rew, Lynn, Deborah Koniak-Griffin, Mary Ann Lewis, Margaret Miles, and Ann O'Sullivan. 2000. "Secondary Data Analysis: New Perspective for Adolescent Research." Nursing Outlook 48:223–229.

Reynolds, Paul Davidson. 1979. Ethical Dilemmas and Social Science Research. San Francisco: Jossey-Bass.

Richards, Thomas J. and Lyn Richards. 1994. "Using Computers in Qualitative Research." Pp. 445–462 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Richardson, Laurel. 1995. "Narrative and Sociology." Pp. 198–221 in Representation in Ethnography, edited by John Van Maanen. Thousand Oaks, CA: SAGE.

Riedel, Marc. 2000. Research Strategies for Secondary Data: A Perspective for Criminology and Criminal Justice. Thousand Oaks, CA: SAGE.

Riessman, Catherine Kohler. 2008. Narrative Methods for the Human Sciences. Thousand Oaks, CA: SAGE.

Rinehart, Jenny K. and Elizabeth A. Yeater. 2011. "A Qualitative Analysis of Sexual Victimization Narratives." Violence Against Women 17(7):925–943.

Ringwalt, Christopher L., Jody M. Greene, Susan T. Ennett, Ronaldo Iachan, Richard R. Clayton, and Carl G. Leukefeld. 1994. Past and Future Directions of the D.A.R.E.

Program: An Evaluation Review. Research Triangle, NC: Research Triangle Institute.

Rives, Norfleet W. Jr. and William J. Serow. 1988. Introduction to Applied Demography: Data Sources and Estimation Techniques. SAGE University Paper Series on Quantitative Applications in the Social Sciences, series No. 07–039. Thousand Oaks, CA: SAGE.

Robertson, David Brian. 1993. “The Return to History and the New Institutionalism in American Political Science.” Social Science History 17:1–36.

Robinson, Karen A. and Steven N. Goodman. 2011. “A Systematic Examination of the Citation of Prior Research in Reports of Randomized, Controlled Trials.” Annals of Internal Medicine 154:50–55.

Rodríguez, Havidán, Joseph Trainor, and Enrico L. Quarantelli. 2006. “Rising to the Challenges of a Catastrophe: The Emergent and Prosocial Behavior Following Hurricane Katrina.” The Annals of the American Academy of Political and Social Science 604:82–101.

Roman, Anthony. 2005. Women’s Health Network Client Survey: Field Report. Unpublished report. Boston: Center for Survey Research, University of Massachusetts.

Rookey, Bryan D., Steve Hanway, and Don A. Dillman. 2008. “Does a Probability-Based Household Panel Benefit From Assignment to Postal Response as an Alternative to Internet-Only?” Public Opinion Quarterly 72:962–984.

Rosen, Lawrence. 1995. “The Creation of the Uniform Crime Report: The Role of Social Science.” Social Science History 19:215–238.

Rosenbach, Margo, Carol Irvin, Angela Merrill, Shanna Shulman, John Czajka, Christopher Trenholm, Susan Williams, So Sasigant Limpa-Amara, and Anna Katz. 2007. National Evaluation of the State Children’s Health Insurance Program: A Decade of Expanding Coverage and Improving Access: Final Report. Cambridge, MA: Mathematica Policy Research.

Rosenbaum, Dennis P. 2007. “Just Say No to D.A.R.E.” Criminology & Public Policy 6:815–824.

Rosenbaum, Dennis P. and Gordon S. Hanson.1998. “Assessing the Effects of School-Based Drug Education: A Six-Year Multi-Level Analysis of Project D.A.R.E.” Journal of Research in Crime and Delinquency 35:381–412.

Rosenberg, Morris. 1968. The Logic of Survey Analysis. New York: Basic Books.

Rosenfeld, Richard. 2004. “The Case of the Unsolved Crime Decline.” Scientific American 290:82–89.

Rosenthal, Elisabeth. 2000. “Rural Flouting of One-Child Policy Undercuts China’s Census.” The New York Times, April 14, p. A10.

Rosenthal, Rob. 1994. Homeless in Paradise: A Map of the Terrain. Philadelphia: Temple University Press.

Ross, Catherine E. 1990. Work, Family, and the Sense of Control: Implications for the Psychological Well-Being of Women and Men. Proposal submitted to the National Science Foundation. Urbana: University of Illinois.

Rossi, Peter H. 1989. Down and Out in America: The Origins of Homelessness. Chicago: University of Chicago Press.

Rossi, Peter H. 1999. “Half Truths With Real Consequences: Journalism, Research, and Public Policy. Three Encounters.” Contemporary Sociology 28:1–5.

Rossi, Peter H. and Howard E. Freeman. 1989. Evaluation: A Systematic Approach, 4th ed. Newbury Park, CA: SAGE.

Rossman, Gretchen B. and Sharon F. Rallis. 1998. Learning in the Field: An Introduction to Qualitative Research. Thousand Oaks, CA: SAGE.

Roth, Dee. 1990. “Homelessness in Ohio: A Statewide Epidemiological Study.” Pp. 145–163 in Homeless in the United States, Vol. 1: State Surveys, edited by Jamshid Momeni. New York: Greenwood.

Rotolo, Thomas and Charles R. Tittle. 2006. “Population Size, Change, and Crime in U.S. Cities.” Journal of Quantitative Criminology 22:341–367.

Rubin, Herbert J. and Irene S. Rubin. 1995. Qualitative Interviewing: The Art of Hearing Data. Thousand Oaks, CA: SAGE.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. Capitalist Development and Democracy. Chicago: University of Chicago Press.

Ruggles, Patricia. 1990. Drawing the Line: Alternative Poverty Measures and Their Implications for Public Policy. Washington, DC: Urban Institute Press.

Sacks, Stanley, Karen McKendrick, George DeLeon, Michael T. French, and Kathryn E. McCollister. 2002. “Benefit-Cost Analysis of a Modified Therapeutic Community for Mentally Ill Chemical Abusers.” Evaluation and Program Planning 25:137–148.

Salisbury, Robert H. 1975. “Research on Political Participation.” American Journal of Political Science 19(May):323–341.

Salmons, Janet. 2012. “Designing and Conducting Research With Online Interviews.” Pp. 1–35 in Cases in Online Interview Research, edited by Janet Salmons. Thousand Oaks, CA: SAGE.

Salmons, Janet E. 2016. Doing Qualitative Research Online. Thousand Oaks, CA: SAGE.

Sampson, Robert J. 1987. “Urban Black Violence: The Effect of Male Joblessness and Family Disruption.” American Journal of Sociology 93(September):348–382.

Sampson, Robert J. and John H. Laub. 1990. “Crime and Deviance Over the Life Course: The Salience of Adult Social Bonds.” American Sociological Review

55(October):609–627.

Sampson, Robert J. and John H. Laub. 1993. "Structural Variations in Juvenile Court Processing: Inequality, the Underclass, and Social Control." Law & Society Review 27(2):285–311.

Sampson, Robert J. and John H. Laub. 1994. "Urban Poverty and the Family Context of Delinquency: A New Look at Structure and Process in a Classic Study." Child Development 65:523–540.

Sampson, Robert J. and Stephen W. Raudenbush. 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." American Journal of Sociology 105:603–651.

Sampson, Robert J. and Stephen W. Raudenbush. 2001. "Disorder in Urban Neighborhoods—Does It Lead to Crime?" In Research in Brief. Washington, DC: National Institute of Justice, U.S. Department of Justice.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." Science 277:918–924.

Sandstrom, Heather, Sarah Benatar, Erica Greenberg, Ian Hill, Embry Howell, Timothy Triplett, Margaret Wilkinson, Todd Franke, Christina Christie, Maria Lourdes Brown, and Jennifer Ho. 2015. Welcome Baby Home Visiting: Findings from the 36-Month Child & Family Survey and 3-Year Longitudinal Results. Urban Institute. Retrieved from http://www.first5la.org/postfiles/files/36-Month%20Outcomes%20Report.pdf

Sayer, Andrew. 2003. "Reductionism in Social Science." Lancaster, England: Department of Sociology, Lancaster University. Retrieved from www.lancaster.ac.uk/fass/sociology/research/publications/papers/sayer-paris1.pdf

Schaeffer, Nora Cate and Stanley Presser. 2003. "The Science of Asking Questions." Annual Review of Sociology 29:65–88.

Schegloff, Emanuel A. 1996. “Issues of Relevance for Discourse Analysis: Contingency in Action, Interaction and Coparticipant Context.” Pp. 3–35 in Computational and Conversational Discourse: Burning Issues—An Interdisciplinary Account, edited by Eduard H. Hovy and Donia R. Scott. New York: Springer.

Schleyer, Titus K. L. and Jane L. Forrest. 2000. “Methods for the Design and Administration of Web-Based Surveys.” Journal of the American Medical Informatics Association 7:418–425.

Schober, Michael F. 1999. “Making Sense of Survey Questions.” Pp. 77–94 in Cognition and Survey Research, edited by Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwartz, Judith M. Tanur, and Roger Tourangeau. New York: Wiley.

Schofield, Janet Ward. 2002. “Increasing the Generalizability of Qualitative Research.” Pp. 171–203 in The Qualitative Researcher’s Companion, edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: SAGE.

Schorr, Lisbeth B. and Daniel Yankelovich. 2000. “In Search of a Gold Standard for Social Programs.” The Boston Globe, February 18, p. A19.

Schreck, Christopher J., Eric A. Steward, and Bonnie S. Fisher. 2006. “Self-Control, Victimization, and Their Influence on Risky Lifestyles: A Longitudinal Analysis Using Panel Data.” Journal of Quantitative Criminology 22:319–340.

Schreer, George E., Saundra Smith, and Kirsten Thomas. 2009. “‘Shopping While Black’: Examining Racial Discrimination in a Retail Setting.” Journal of Applied Social Psychology 39:1432–1444.

Schuman, Howard and Stanley Presser. 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Schutt, Russell K. 1986. Organization in a Changing Environment. Albany: State University of New York Press.

Schutt, Russell K. (with the assistance of Tatjana Meschede). 1992. The Perspectives of DMH Shelter Staff: Their Clients, Their Jobs, Their Shelters and the Service System. Unpublished report to the Metro Boston Region of the Massachusetts Department of Mental Health. Boston: Department of Sociology, University of Massachusetts.

Schutt, Russell K. 2011a. “Evaluation of the Coordinated Care Program.” Proposal to Institutional Review Board for the Protection of Human Subjects, University of Massachusetts Boston.

Schutt, Russell K. 2011b. Homelessness, Housing, and Mental Illness. Cambridge, MA: Harvard University Press.

Schutt, Russell K. and W. Dale Dannefer. 1988. “Detention Decisions in Juvenile Cases: JINS, JDs and Gender.” Law & Society Review 22(3):509–520.

Schutt, Russell K., Xiaogang Deng, Gerald R. Garrett, Stephanie Hartwell, Sylvia Mignon, Joseph Bebo, Matthew O’Neill, Mary Aruda, Pat Duynstee, Pam DiNapoli, and Helen Reiskin. 1996. Substance Use and Abuse Among UMass Boston Students. Boston: Department of Sociology, University of Massachusetts. Unpublished report.

Schutt, Russell K. and Jacqueline Fawcett. 2005. Case Management in the Women’s Health Network. Boston: University of Massachusetts. Unpublished report.

Schutt, Russell K. and M. L. Fennell. 1992. “Shelter Staff Satisfaction With Services, the Service Network and Their Jobs.” Current Research on Occupations and Professions 7:177–200.

Schutt, Russell K. and Stephen M. Goldfinger. 1996. “Housing Preferences and Perceptions of Health and Functioning Among Homeless Mentally Ill Persons.” Psychiatric Services 47:381–386.

Schutt, Russell K., Stephen M. Goldfinger, and Walter E. Penk. 1997. “Satisfaction With Residence and With Life: When Homeless Mentally Ill Persons Are Housed.” Evaluation and Program Planning 20(2):185–194.

Schutt, Russell K., Suzanne Gunston, and John O'Brien. 1992. "The Impact of AIDS Prevention Efforts on AIDS Knowledge and Behavior Among Sheltered Homeless Adults." Sociological Practice Review 3(1):1–7.

Schutt, Russell K., Richard L. Hough, Stephen M. Goldfinger, Anthony F. Lehman, David L. Shern, Elie S. Valencia, and Patricia A. Wood. 2009. "Lessening Homelessness Among Persons With Mental Illness: A Comparison of Five Randomized Treatment Trials." Asian Journal of Psychiatry 2:100–105. Retrieved from www.ncbi.nlm.nih.gov/pmc/articles/PMC2788308/pdf/nihms147743.pdf

Schwandt, Thomas A. 1994. "Constructivist, Interpretivist Approaches to Human Inquiry." Pp. 118–137 in Handbook of Qualitative Research, edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Schwartz, John. 2005. "Myths Run Wild in Blog Tsunami Debate." The New York Times, January 3, p. A9.

Schwarz, Norbert. 2010. "Measurement as Cooperative Communication: What Research Participants Learn From Questionnaires." Pp. 43–59 in The SAGE Handbook of Measurement, edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: SAGE.

Scott, Janny. 2001. "A Nation by the Numbers, Smudged." The New York Times, July 1, pp. 21, 22.

Scott, John. 2013. Social Network Analysis, 3rd ed. Thousand Oaks, CA: SAGE.

Scriven, Michael. 1972a. "The Methodology of Evaluation." Pp. 123–136 in Evaluating Action Programs: Readings in Social Action and Education, edited by Carol H. Weiss. Boston: Allyn & Bacon.

Scriven, Michael. 1972b. "Prose and Cons About Goal-Free Evaluation." Evaluation Comment 3:1–7.

Scull, Andrew T. 1988. “Deviance and Social Control.” Pp. 667–693 in Handbook of Sociology, edited by Neil J. Smelser. Newbury Park, CA: SAGE.

Sechrest, Lee and Souraya Sidani. 1995. “Quantitative and Qualitative Methods: Is There an Alternative?” Evaluation and Program Planning 18:77–87.

Sechrist, Stacy M. and John D. Weil. 2017. “Assessing the Impact of a Focused Deterrence Strategy to Combat Intimate Partner Domestic Violence.” Violence Against Women. Online first.

Selm, Martine Van and Nicholas W. Jankowski. 2006. “Conducting Online Surveys.” Quality & Quantity 40:435–456.

Selwitz, Ada Sue, Norma Epley, and Janelle Erickson. 2013. “Basic Institutional Review Board (IRB): Regulations and Review Process.” Miami, FL: Collaborative Institutional Training Initiative at the University of Miami. Retrieved from https://www.citiprogram.org/

Shadish, William R., Thomas D. Cook, and Laura C. Leviton (Eds.). 1991. Foundations of Program Evaluation: Theories of Practice. Thousand Oaks, CA: SAGE.

Sharma, Divya. 2009. “Research Ethics and Sensitive Behaviors: Underground Economy.” Pp. 426–441 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Shepherd, Jane, David Hill, Joel Bristor, and Pat Montalvan. 1996. “Converting an Ongoing Health Study to CAPI: Findings From the National Health and Nutrition Study.” Pp. 159–164 in Health Survey Research Methods Conference Proceedings, edited by Richard B. Warnecke. Hyattsville, MD: U.S. Department of Health and Human Services.

Sherman, Lawrence W. 1992. Policing Domestic Violence: Experiments and Dilemmas. New York: Free Press.

Sherman, Lawrence W. 1993. “Implications of a Failure to Read the Literature.” American Sociological Review 58:888–889.

Sherman, Lawrence W. and Richard A. Berk. 1984. “The Specific Deterrent Effects of Arrest for Domestic Assault.” American Sociological Review 49:261–272.

Sherman, Lawrence W. and Heather M. Harris. 2013. “Increased Homicide Victimization of Suspects Arrested for Domestic Assault: A 23-Year Follow-Up of the Milwaukee Domestic Violence Experiment (MilDVE). Journal of Experimental Criminology 9:491–514.

Sherman, Lawrence W. and Douglas A. Smith, with Janell D. Schmidt and Dennis P. Rogan. 1992. “Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence.” American Sociological Review 57:680–690.

Shermer, Michael. 1997. Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time. New York: W. H. Freeman.

Shrout, Patrick E. 2011. “Integrating Causal Analysis into Psychopathology Research.” Pp. 3–24 in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures, edited by Patrick E. Shrout, Katherine M. Keyes, and Katherine Ornstein. New York: Oxford University Press.

Shrum, Wesley and Greg Scott. 2017. Video Ethnography in Practice: Planning, Shooting, and Editing for Social Analysis. Thousand Oaks, CA: SAGE.

Sieber, Joan E. 1992. Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards. Thousand Oaks, CA: SAGE.

Sieber, Joan E. and Martin B. Tolich. 2013. Planning Ethically Responsible Research, 2nd ed. Thousand Oaks, CA: SAGE.

Sin, Chih Hoong. 2005. “Seeking Informed Consent: Reflections on Research Practice.” Sociology 39:277–294.

Sjoberg, Gideon (Ed.). 1967. Ethics, Politics, and Social Research. Cambridge, MA: Schenkman.

Skinner, Harvey A. and Wen-Jenn Sheu. 1982. “Reliability of Alcohol Use Indices: The Lifetime Drinking History and the MAST.” Journal of Studies on Alcohol 43(11):1157–1170.

Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. New York: Cambridge University Press.

Skocpol, Theda. 1984. “Emerging Agendas and Recurrent Strategies in Historical Sociology.” Pp. 356–391 in Vision and Method in Historical Sociology, edited by Theda Skocpol. New York: Cambridge University Press.

Skocpol, Theda and Margaret Somers. 1979. “The Uses of Comparative History in Macrosocial Inquiry.” Pp. 72–95 in Social Revolutions in the Modern World, edited by Theda Skocpol. New York: Cambridge University Press.

Slater, Mel, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V. Sanchez-Vives. 2006. “A Virtual Reprise of the Stanley Milgram Obedience Experiments.” PLoS ONE, 1(1).

Sloboda, Zili, Richard C. Stephens, Peggy C. Stephens, Scott F. Grey, Brent Teasdale, Richard D. Hawthorne, Joseph Williams, and Jesse F. Marquette. 2009. “The Adolescent Substance Abuse Prevention Study: A Randomized Field Trial of a Universal Substance Abuse Prevention Program.” Drug and Alcohol Dependence 102(1–3):1–10.

Smith, Erica L. and Alexia Cooper. 2013. Homicide in the U.S. Known to Law Enforcement, 2011. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Smith, Tom W. 1984. “Nonattitudes: A Review and Evaluation.” Pp. 215–255 in Surveying Subjective Phenomena, vol. 2, edited by Charles F. Turner and Elizabeth Martin. New York: Russell Sage Foundation.

Smith, Tom W. 1995. "Trends in Nonresponse Rates." International Journal of Public Opinion Research 7: 157–171.

Smith, Tom W., Michael Davern, Jeremy Freese, and Michael Hout. 2017. General Social Surveys, 1972–2016. [machine-readable data file]. Principal Investigator, Tom W. Smith; Co-Principal Investigators, Peter V. Marsden and Michael Hout, NORC ed. Chicago: NORC. 1 data file (62,466 logical records) and 1 codebook (3,689 pp).

Smithson, Janet. 2008. "Focus Groups." Pp. 357–370 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2004. How Visual Grouping Influences Answers to Internet Surveys. Extended version of paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ, May 13. Retrieved from http://survey.sesrc.wsu.edu/dillman/papers.htm

Snipp, C. Matthew. 2003. "Racial Measurement in the American Census: Past Practices and Implications for the Future." Annual Review of Sociology 29:563–588.

Snow, David L., Jacob Kraemer Tebes, and Michael W. Arthur. 1992. "Panel Attrition and External Validity in Adolescent Substance Use Research." Journal of Consulting and Clinical Psychology 60:804–807.

Sobell, Linda C., Mark B. Sobell, Diane M. Riley, Reinhard Schuller, D. Sigfrido Pavan, Anthony Cancilla, Felix Klajner, and Gloria I. Leo. 1988. "The Reliability of Alcohol Abusers' Self-Reports of Drinking and Life Events That Occurred in the Distant Past." Journal of Studies on Alcohol 49(2):225–232.

Social Trends Institute. 2017. The Sustainable Demographic Dividend: What Do Marriage & Fertility Have to Do With the Economy? Retrieved from http://sustaindemographicdividend.org/articles/international-family-indicators/global-family-structure

Sorowski, Piotr, Emanuel Kulczycki, Agnieszka Sorokowska, and Katarzyna Pisanski. 2017. “Predatory Journals Recruit Fake Editor.” Nature 543 (March 23):481–483.

Sosin, Michael R., Paul Colson, and Susan Grossman. 1988. Homelessness in Chicago: Poverty and Pathology, Social Institutions and Social Change. Chicago: Chicago Community Trust.

Speiglman, Richard and Patricia Spear. 2009. “The Role of Institutional Review Boards: Ethics: Now You See Them, Now You Don’t.” Pp. 121–134 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Spretnak, Charlene. 1991. States of Grace: The Recovery of Meaning in the Postmodern Age. New York: HarperCollins.

St. Jean, Peter K. B. 2007. Pockets of Crime: Broken Windows, Collective Efficacy, and the Criminal Point of View. Chicago: University of Chicago Press.

St. Pierre, Robert G. and Peter H. Rossi. 2006. “Randomize Groups, Not Individuals: A Strategy for Improving Early Childhood Programs.” Evaluation Review 30:656–685.

Stake, Robert E. 1995. The Art of Case Study Research. Thousand Oaks, CA: SAGE.

Stake, Robert and Fazal Rizvi. 2009. “Research Ethics in Transnational Spaces.” Pp. 521–536 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

Stephen, Jason M., Michael F. Young, and Thomas Calabrese. 2007. “Does Moral Judgment Go Offline When Students Are Online? A Comparative Analysis of Undergraduates’ Beliefs and Behaviors Related to Conventional and Digital Cheating.” Ethics & Behavior 17:233–254.

Stewart, David W. and Michael A. Kamins. 1993. Secondary Research: Information Sources and Methods, 2nd ed. Newbury Park, CA: SAGE.

Stewart, Nikita. 2017. “As More Opt for Streets, City Offers a Place to Go.” The New York Times, July 19, p. A20.

Stille, Alexander. 2000. “A Happiness Index With a Long Reach: Beyond G.N.P. to Subtler Measures.” The New York Times, May 20, pp. A17, A19.

Stokoe, Elizabeth. 2006. “On Ethnomethodology, Feminism, and the Analysis of Categorical Reference to Gender in Talk-in-Interaction.” The Sociological Review 54:467–494.

Strahan, Robert and Kathleen Carrese Gerbasi. 1972. “Short, Homogeneous Versions of the Marlowe-Crowne Social Desirability Scale.” Journal of Clinical Psychology 28:191–193.

Strauss, Anselm L. and Juliette Corbin. 1990. The Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park, CA: SAGE.

Strunk, William Jr. and E. B. White. 2000. The Elements of Style, 4th ed. New York: Allyn & Bacon.

Sudman, Seymour. 1976. Applied Sampling. New York: Academic Press.

Sue, Valerie M. and Lois A. Ritter. 2012. Conducting Online Surveys, 2nd ed. Thousand Oaks, CA: SAGE.

Sulkunen, Pekka. 2008. “Social Research and Social Practice in Post-Positivist Society.” Pp. 68–80 in The SAGE Handbook of Social Research Methods, edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: SAGE.

Sunderland, Antonia. 2005. Children, Families and Welfare Reform: A Three-City Study. Princeton, NJ: The Robert Wood Johnson Foundation.

SurveyMonkey. n.d. “Engaged Survey Respondents, Honest Feedback.” Retrieved from https://www.surveymonkey.com/mp/audience/our-survey-respondents

"Survey on Adultery: 'I Do' Means 'I Don't.'" 1993. The New York Times, October 19, p. A20.

Survey Research Laboratory. 2008. List of Academic and Not-for-Profit Survey Research Organizations (LANSRO). Chicago: College of Urban Planning and Public Affairs, University of Illinois at Chicago. Retrieved from www.srl.uic.edu/LANSRO.doc

Swarns, Rachel L. 2004. "Hispanics Debate Racial Grouping by Census." The New York Times, October 24, pp. A1, A18.

Tavernise, Sabrina. 2002. "How Many Russians? Let Us Weigh the Count, Cooperation or No." The New York Times, October 10, p. A13.

Tavory, Iddo and Stefan Timmermans. 2014. Abductive Analysis: Theorizing Qualitative Research. Chicago: University of Chicago Press.

Taylor, Charles Lewis and David A. Jodice. 1986. World Handbook of Political and Social Indicators III: 1948–1982. File available from the Inter-university Consortium for Political and Social Research (ICPSR), Study #7761.

Taylor, Jerry. 1999. "DARE Gets Updated in Some Area Schools, Others Drop Program." The Boston Sunday Globe, May 16, pp. 1, 11.

Testa, Maria, Jennifer A. Livingston, and Carol VanZile-Tamsen. 2011. "Advancing the Study of Violence Against Women Using Mixed Methods: Integrating Qualitative Methods Into a Quantitative Research Program." Violence Against Women 17(2): 236–250.

Thomas, Neil. 2005. "Disaster Center Researchers to Study Katrina Response." UDaily, September 15. Retrieved from www.udel.edu/PR/UDaily/2005/mar/DRC091505.html

Thorne, Barrie. 1993. Gender Play: Girls and Boys in School. New Brunswick, NJ: Rutgers University Press.

Thorsen, Maggie L. and Jennifer Pearce-Morris. 2016. “Adolescent Mental Health and Dating in Young Adulthood.” Society and Mental Health 6(3):223–245.

Thrasher, Frederic M. 1927. The Gang: A Study of 1,313 Gangs in Chicago. Chicago: University of Chicago Press.

Tigges, Beth Baldwin. 2003. “Parental Consent and Adolescent Risk Behavior Research.” Journal of Nursing Scholarship 35:283–289.

Tight, Malcolm. 2017. Understanding Case Study Research: Small-Scale Research With Meaning. Thousand Oaks, CA: SAGE.

Timmer, Doug A., D. Stanley Eitzen, and Kathryn D. Talley. 1993. Paths to Homelessness: Extreme Poverty and the Urban Housing Crisis. Boulder, CO: Westview Press.

Tinkler, Penny. 2013. Using Photographs in Social and Historical Research. Thousand Oaks, CA: SAGE.

Tjaden, Patricia and Nancy Thoennes. 2000. Extent, Nature, and Consequences of Intimate Partner Violence: Findings From the National Violence Against Women Survey, NCJ 181867. Washington, DC: Office of Justice Programs, National Institute of Justice and the Centers for Disease Control and Prevention.

Toby, Jackson. 1957. “Social Disorganization and Stake in Conformity: Complementary Factors in the Predatory Behavior of Hoodlums.” Journal of Criminal Law, Criminology and Police Science 48:12–17.

Toepoel, Vera. 2016. Doing Surveys Online. Thousand Oaks, CA: SAGE.

Toepoel, Vera, Alerk Amin, and Peter Lugtig. n.d. “The Effects of Adding a Mobile-Compatible Design to the American Life Panel.” RAND Labor and Population, American Life Panel. Retrieved from http://slideplayer.com/slide/9015836/

Toppo, Greg. 2002. “Antidrug Program Backed by Study.” The Boston Globe, October 29, p. A10.

Tourangeau, Roger. 1999. “Context Effects.” Pp. 111–132 in Cognition and Survey Research, edited by Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwartz, Judith M. Tanur, and Roger Tourangeau. New York: Wiley.

Tourangeau, Roger. 2004. “Survey Research and Societal Change.” Annual Review of Psychology 55:775–801.

Tourangeau, Roger, Frederick G. Conrad, and Mick P. Couper. 2012. The Science of Web Surveys. Oxford: Oxford University Press.

Townsend, Meg, Dana Hunt, Caity Baxter, and Peter Finn. 2005. Interim Report: Evaluability Assessment of the President’s Family Justice Center Initiative. Cambridge, MA: Abt Associates. Retrieved from www.ncjrs.gov/pdffiles1/nij/grants/212278.pdf

Tufte, Edward R. 1983. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Turabian, Kate L. 2007. A Manual for Writers of Term Papers, Theses, and Dissertations, 7th ed. Chicago: University of Chicago Press.

Turkle, Sherry. 2011. Alone Together: Why We Expect More From Technology and Less From Each Other. New York: Basic Books.

Turkle, Sherry. 2015. Reclaiming Conversation: The Power of Talk in a Digital Age. New York: Penguin Books.

Turner, Charles F. and Elizabeth Martin (Eds.). 1984. Surveying Subjective Phenomena, vols. 1 and 2. New York: Russell Sage Foundation.

Turner, Jonathan H., Leonard Beeghley, and Charles H. Powers. 1995. The Emergence of Sociological Theory, 3rd ed. Belmont, CA: Wadsworth.

Tyler, Tom R. 1990. “The Social Psychology of Authority: Why Do People Obey an Order to Harm Others?” Law & Society Review 24:1089–1102.

Uchitelle, Louis. 1997. “Measuring Inflation: Can’t Do It, Can’t Stop Trying.” The New York Times, March 16, p. 4.

Uchitelle, Louis. 1999. “Devising New Math to Define Poverty.” The New York Times, October 16, pp. A1, A14.

UCLA Center for Communication Policy. 2001. The UCLA Internet Report 2001: Surveying the Digital Future. Los Angeles: UCLA Center for Communication Policy.

UCLA Center for Communication Policy. 2003. The UCLA Internet Report: Surveying the Digital Future, Year Three. Los Angeles: UCLA Center for Communication Policy. Retrieved from www.digitalcenter.org/pdf/InternetReportYearThree.pdf

U.S. Bureau of Economic Analysis. 2004. Customer Satisfaction Survey Report, FY 2004. Washington, DC: U.S Department of Commerce.

U.S. Bureau of Labor Statistics, Department of Labor. 1991. Major Programs of the Bureau of Labor Statistics. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997a. Employment and Earnings. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997b. Handbook of Methods. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 2014. National Longitudinal Survey of Youth: Children and Young Adults. Retrieved from https://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/questionnaires

U.S. Census Bureau. 1981. Section 1, Vital Statistics. Statistical Abstract of the United States, 1981, 102nd ed. Washington, DC: U.S. Department of Commerce, Census Bureau.

U.S. Census Bureau. 1994. Census Catalog and Guide, 1994. Washington, DC: U.S. Bureau of the Census.

U.S. Census Bureau. 1999. United States Census 2000, Updated Summary: Census 2000 Operational Plan. Washington, DC: U.S. Department of Commerce, Bureau of the Census, February.

U.S. Census Bureau. 2001. "Statement by William G. Barron Jr. on the Current Status of Results of Census 2000 Accuracy and Coverage Evaluation Survey." United States Department of Commerce News, July 13. Retrieved from www.census.gov/PressRelease/www/2001/cb01cs06.html

U.S. Census Bureau. 2006. Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials. Washington, DC: U.S. Census Bureau, Census Advisory Committees. Retrieved from www.census.gov/cac/www/007585.html

U.S. Census Bureau. 2010a. "$1.6 Billion in 2010 Census Savings Returned." United States Department of Commerce News, August 10. Retrieved from http://2010.census.gov/news/releases/operations/

U.S. Census Bureau. 2010b. "Door-to-Door Visits Begin for 2010 Census." United States Department of Commerce News, April 30. Retrieved from http://2010.census.gov/news/releases/operations/

U.S. Census Bureau. 2010c. "2010 Census Forms Arrive in 120 Million Mailboxes Across Nation." United States Department of Commerce News, March 15. Retrieved from http://2010.census.gov/news/releases/operations/

U.S. Census Bureau. 2010d. Questions and Answers: Real People, Real Questions, Real Answers. Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Retrieved from http://2010.census.gov/2010census/about/answers.php

U.S. Census Bureau. 2013. “Census Bureau News—Extended Measures of Well-Being: Living Conditions in the United States: 2011.” Washington, DC: U.S. Census Bureau. Retrieved from www.prnewswire.com/news-releases/census-bureau-news----extended-measures-of-well-being-living-conditions-in-the-united-states-2011-222530671.html

U.S. Census Bureau. 2014. “Foreign-Born: 2014 Current Population Survey Detailed Tables.” Retrieved from https://www.census.gov/data/tables/2014/demo/foreign-born/cps-2014.html

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. 1995. Client- Level Evaluation Procedure Manual. Washington, DC: U.S. Department of Health and Human Services.

U.S. National Library of Medicine. n.d. “ClinicalTrials.gov: History, Policies, and Law.” Retrieved from https://clinicaltrials.gov/ct2/about-site/history

U.S. Office of Management and Budget. 2002. Government and Performance Results Act of 1993. Washington, DC: U.S. Office of Management and Budget, Executive Office of the President.

Vaessen, Martin. 1993. “Evaluation of Population Data: Errors and Deficiencies.” Pp. 4-1–4-69 in Readings in Population Research Methodology, Vol. 1, Basic Tools, edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United Nations Population Fund.

Vaillant, George E. 1995. The Natural History of Alcoholism Revisited. Cambridge, MA: Harvard University Press.

van de Vijver, Fons and Kwok Leung. 1997. Methods and Data Analysis for Cross-Cultural Research. Thousand Oaks, CA: SAGE.

Van Hoye, Greet and Filip Lievens. 2003. “The Effects of Sexual Orientation on Hirability Ratings: An Experimental Study.” Journal of Business and Psychology 18:15–30.

Van Maanen, John. 1982. “Fieldwork on the Beat.” Pp. 103–151 in Varieties of Qualitative Research, edited by John Van Maanen, James M. Dabbs Jr., and Robert R. Faulkner. Beverly Hills, CA: SAGE.

Van Maanen, John. 1995. “An End to Innocence: The Ethnography of Ethnography.” Pp. 1–35 in Representation in Ethnography, edited by John Van Maanen. Thousand Oaks, CA: SAGE.

Vargas, Roberto B., Gery W. Ryan, Catherine A. Jackson, Rian Rodriquez, and Harold P. Freeman. 2008. “Characteristics of the Original Patient Navigation Programs to Reduce Disparities in the Diagnosis and Treatment of Breast Cancer.” Cancer 113:426–433.

Vega, Tanzina. 2014. “To Measure More Diverse America, Solution May Be in Census Questions.” The New York Times, July 2, pp. A12, A16.

Venkatapuram, Sridhar. 2013. “Subjective Wellbeing: A Primer for Poverty Analysts.” Journal of Poverty and Social Justice 21:5–17.

Venkatesh, Sudhir. 2008. Gang Leader for a Day: A Rogue Sociologist Takes to the Streets. New York: Penguin.

Verba, Sidney and Norman Nie. 1972. Political Participation: Political Democracy and Social Equality. New York: Harper & Row.

Verba, Sidney, Norman Nie, and Jae-On Kim. 1978. Participation and Political Equality: A Seven-Nation Comparison. New York: Cambridge University Press.

Vernez, Georges, M. Audrey Burnam, Elizabeth A. McGlynn, Sally Trude, and Brian S. Mittman. 1988. Review of California’s Program for the Homeless Mentally Disabled. Santa Monica, CA: RAND.

Vidich, Arthur J. and Stanford M. Lyman. 2004. “Qualitative Methods: Their History in Sociology and Anthropology.” Pp. 27–84 in The Handbook of Qualitative Research, 2nd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Villarreal, Andrés. 2010. “Stratification by Skin Color in Contemporary Mexico.” American Sociological Review 75:652–678.

Vincus, Amy A., Chris Ringwalt, Melissa S. Harris, and Stephen R. Shamblen. 2010. “A Short-Term, Quasi-Experimental Evaluation of D.A.R.E.’s Revised Elementary School Curriculum.” Journal of Drug Education 40:37–49.

Viswanathan, Madhu. 2005. Measurement Error and Research Design. Thousand Oaks, CA: SAGE.

Wageman, Ruth. 1995. “Interdependence and Group Effectiveness.” Administrative Science Quarterly 40:145–180.

Wagner, William E. III. 2011. Using IBM SPSS Statistics for Social Statistics and Research Methods, 3rd ed. Thousand Oaks, CA: SAGE.

Waksberg, Joseph. 1978. “Sampling Methods for Random Digit Dialing.” Journal of the American Statistical Association 73:40–46.

Walker, Michael E., Stanley Wasserman, and Barry Wellman. 1993. “Statistical Models for Social Support Networks.” Sociological Methods & Research 22: 71–98.

Walker, Robert, Mark Tomlinson, and Glenn Williams. 2010. “The Problem With Poverty: Definition, Measurement and Interpretation.” Pp. 353–376 in The SAGE Handbook of Measurement, edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: SAGE.

Wallace, Walter L. 1983. Principles of Scientific Sociology. New York: Aldine.

Wallgren, Anders, Britt Wallgren, Rolf Persson, Ulf Jorner, and Jan-Aage Haaland. 1996. Graphing Statistics and Data: Creating Better Charts. Thousand Oaks, CA: SAGE.

Walters, Pamela Barnhouse, David R. James, and Holly J. McCammon. 1997. “Citizenship and Public Schools: Accounting for Racial Inequality in Education for the Pre- and Post-Disfranchisement South.” American Sociological Review 62:34–52.

Ward, Lester Frank. 1897. Dynamic Sociology, 2nd ed. New York: D. Appleton.

Warner, W. Lloyd, Marchia Meeker, and Kenneth Eells. 1960. Social Class in America: A Manual of Procedure for the Measurement of Social Status. New York: Harper & Row.

Warr, Mark. 1995. “The Polls—Poll Trends: Public Opinion on Crime and Punishment.” Public Opinion Quarterly 59:296–310.

Watson, Charles G., Curt Tilleskjor, E. A. Hoodecheck-Schow, John Pucel, and Lyle Jacobs. 1984. “Do Alcoholics Give Valid Self-Reports?” Journal of Studies on Alcohol 45(4):344–348.

Weatherby, Norman L., Richard Needle, Helen Cesari, Robert Booth, Clyde B. McCoy, John K. Watters, Mark Williams, and Dale D. Chitwood. 1994. “Validity of Self-Reported Drug Use Among Injection Drug Users and Crack Cocaine Users Recruited Through Street Outreach.” Evaluation and Program Planning 17(4):347–355.

Weaver-Hightower, Marcus B. 2013. “A Mixed Methods Approach for Identifying Influence on Public Policy.” Journal of Mixed Methods Research 20:1–24.

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 2000. Unobtrusive Measures, rev. ed. Thousand Oaks, CA: SAGE.

Weber, Max. 1949. The Methodology of the Social Sciences. Translated and edited by Edward A. Shils and Henry A. Finch. New York: Free Press.

Weber, Robert Philip. 1990. Basic Content Analysis, 2nd ed. Newbury Park, CA: SAGE.

Wechsler, Henry, Jae Eun Lee, Meichun Kuo, Mark Seibring, Toben F. Nelson, and Hang Lee. 2002. “Trends in College Binge Drinking During a Period of Increased Prevention Efforts.” Journal of American College Health 50:203–217.

Weinberg, Darin. 2000. “‘Out There’: The Ecology of Addiction in Drug Abuse Treatment Discourse.” Social Problems 47:606–621.

Weisburd, David, Cynthia M. Lum, and Anthony Petrosino. 2001. “Does Research Design Affect Study Outcomes in Criminal Justice?” Annals of the American Academy of Political and Social Science 578:50–70.

Weiss, Carol H. 1993. “Where Politics and Evaluation Research Meet.” Evaluation Practice 14: 93–106.

Wellman, Barry. 2004. “Connecting Communities: On and Offline.” Contexts 3:22–28.

Wellman, Barry, Anabel Quan Haase, James Witte, and Keith Hampton. 2001. “Does the Internet Increase, Decrease, or Supplement Social Capital? Social Networks, Participation, and Community Commitment.” American Behavioral Scientist, 45:436–455.

Wellman, Barry and Keith Hampton. 1999. “Living Networked in a Wired World.” Comparative Sociology 28:1–12.

Wellman, Barry and Scot Wortley. 1990. “Different Strokes From Different Folks: Community Ties and Social Support.” American Journal of Sociology 96:558–588.

Wells, William, Michael R. Cavanaugh, Jeffrey A. Bouffard, and Matt R. Nobles. 2012. “Non-Response Bias With a Web-Based Survey of College Students: Differences From a Classroom Survey About Carrying Concealed Handguns.” Journal of Quantitative Criminology 28:455–476.

Wengraf, Tom. 2001. Qualitative Research Interviewing: Biographic Narrative and Semi-structured Methods. Thousand Oaks, CA: SAGE.

West, Steven L. and Ken K. O’Neal. 2004. “Project D.A.R.E. Outcome Effectiveness Revisited.” American Journal of Public Health 94:1027–1029.

Wheeler, Peter M. 1995. Social Security Programs Throughout the World—1995. Research Report #64, SSA Publication No. 13–11805. Washington, DC: Office of Research and Statistics, Social Security Administration.

White, Aaron M., Courtney L. Kraus, Lindsey A. McCracken, and H. Scott Swartzwelder. 2003. “Do College Students Drink More Than They Think? Use of a Free-Pour Paradigm to Determine How College Students Define Standard Drinks.” Alcoholism: Clinical and Experimental Research 27:1750–1756.

White, Michael J. 1993. “Measurement of Population Size, Composition, and Distribution.” Pp. 1-1–1-29 in Readings in Population Research Methodology, Vol. 1, Basic Tools, edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United Nations Population Fund.

Whyte, William Foote. 1955. Street Corner Society, 2nd ed. Chicago: University of Chicago Press.

Whyte, William Foote. 1991. Social Theory for Social Action: How Individuals and Organizations Learn to Change. Newbury Park, CA: SAGE.

Wiles, Rose, Amanda Coffey, Judy Robinson, and Sue Heath. 2012. “Anonymisation and Visual Images: Issues of Respect, ‘Voice’ and Protection.” International Journal of Social Research Methodology 15:41–53.

Willer, David and Henry A. Walker. 2007. Building Experiments: Testing Social Theory. Stanford, CA: Stanford University Press.

Williams, Kirk R. and Richard Hawkins. 1986. “Perceptual Research on General Deterrence: A Critical Review.” Law and Society Review 20:545–572.

Wilson, Reid. 2017. “New Report Finds That Voter Turnout in 2016 Topped 2012.” The

Hill, March 16. Retrieved from http://thehill.com/homenews/state-watch/324206-new-report-finds-that-voter-turnout-in-2016-topped-2012

Wilson, William Julius. 1998. “Engaging Publics in Sociological Dialogue Through the Media.” Contemporary Sociology 27:435–438.

Wines, Michael. 2017. “Accuracy at Risk as Census Shifts to Online Count.” The New York Times, December 10:1, 20.

Witkin, Belle Ruth and James W. Altschuld. 1995. Planning and Conducting Needs Assessments: A Practical Guide. Thousand Oaks, CA: SAGE.

Witte, James C., Lisa M. Amoroso, and Philip E. N. Howard. 2000. “Research Methodology: Method and Representation in Internet-Based Survey Tools—Mobility, Community, and Cultural Identity in Survey 2000.” Social Science Computer Review 18:179–195.

Wolcott, Harry F. 1995. The Art of Fieldwork. Walnut Creek, CA: AltaMira Press.

Wolf, Amanda, David Turner, and Kathleen Toms. 2009. “Ethical Perspectives in Program Evaluation.” Pp. 170–184 in The Handbook of Social Research Ethics, edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: SAGE.

World Health Organization (WHO). 2005. WHO Multi-Country Study on Women’s Health and Domestic Violence Against Women: Summary Report of Initial Results on Prevalence, Health Outcomes and Women’s Responses. Geneva: World Health Organization.

World Health Organization (WHO). 2013. “Management of Substance Abuse.” Retrieved from www.who.int/substance_abuse/facts/alcohol/en

World Health Organization (WHO). n.d. Process of Translation and Adaptation of Instruments. Retrieved from www.who.int/substance_abuse/research_tools/translation/en

Wunsch, Guillaume J. and Marc G. Termote. 1978. Introduction to Demographic Analysis: Principles and Methods. New York: Plenum Press.

Xu, Yili, Mora L. Fiedler, and Karl H. Flaming. 2005. "Citizens' Judgment Discovering the Impact of Community Policing: The Broken Windows Thesis, Collective Efficacy, and Citizens' Judgment." Journal of Research in Crime and Delinquency 42:147–186.

Yang, Song, Franziska B. Keller, and Lu Zheng. 2017. Social Network Analysis: Methods and Examples. Thousand Oaks, CA: SAGE.

Young, Robert L. 1992. "Religious Orientation, Race and Support for the Death Penalty." Journal for the Scientific Study of Religion 31:76–87.

Zaret, David. 1996. "Petitions and the 'Invention' of Public Opinion in the English Revolution." American Journal of Sociology 101:1497–1555.

Zhang, Sarah. 2015. "The Pitfalls of Using Google Ngram to Study Language." Wired, October 12. Retrieved from https://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram/

Zhong, Juan and Jeffrey J. Arnett. 2014. "Conceptions of Adulthood Among Migrant Women Workers in China." International Journal of Behavioral Development 38:255–265.

Zielbauer, Paul. 2000. "2 Cities Lag Far Behind the U.S. in Heeding the Call of the Census." The New York Times, April 21, p. A21.

Zimbardo, Philip G. 2004. "A Situationist Perspective on the Psychology of Evil: Understanding How Good People Are Transformed Into Perpetrators." Pp. 21–50 in The Social Psychology of Good and Evil: Understanding Our Capacity for Kindness and Cruelty, edited by Arthur G. Miller. New York: Guilford Press.

Zimbardo, Philip. 2007. The Lucifer Effect: Understanding How Good People Turn Evil. New York: Random House.

Zitner, Aaron. 1996. “A Cloudy Gaze Into a Crystal Ball.” The New York Times, September 26, pp. D1, D14.

Index

Exhibits are indicated by “e” following the page number.

Abductive analysis, 430–431
Abma, Tineke, 508
Abortion, 272, 326, 326e
Abrutyn, Seth, 624
Abstract and concrete concepts, 110–115
Abt Associates, 495
Academic motivations for social research, 13
Acceptance, 370–371
Access to Community Care and Effective Services and Supports (ACCESS) program, 120
Acharya, Arnab, 561
Action research, 508
Action tales, 435
Adaptive research design, 371–372
Add Health study, 528
Addiction Severity Index (ASI), 134–135
Advocacy and intervention, 457
African American Health Neighborhood Assessment Scale (AAH NAS), 138, 139e
Aggregate matching, 232–233
Aggression and Violent Behavior, 39
Agreement bias, 265–267
Airbnb, 232
Akyüz, Aygül, 39
Alcohol Abstinence Self-Efficacy Scale (AASE), 271, 272e
Alcoholics Anonymous (AA), 437
Alcohol use disorder (AUD) and intoxication, 103–106
    aggression and, 204–206, 204e, 205e
    content validity and, 133
    contextual effect and, 211
    criterion validity and, 133–134
    difference between reliability and validity in, 140e
    face validity and, 132–133
    idiographic explanations of, 202–203
    survey questions clustered in index and, 270–271
    test-retest reliability and, 136
Alfred, Randall, 380, 381

Allport, Gordon, 105
Alone Together: Why We Expect More From Technology and Less From Each Other, 418
Alternate-forms reliability, 136
Amato, Paul, 137
Ambwani, Suman, 575, 578, 579, 583
American Association of Public Opinion Research (AAPOR), 327
American Community Survey, 527
American Enterprise Institute, 107
American Journal of Sociology, 36
American Medical Association (AMA), 70
American Psychologist, 73–74
American Sociological Association (ASA), 21, 404
    Code of Ethics, 72, 246–247, 612–613
American Sociological Review, 32, 36, 124
Analytic historical sociology, 568
Anderson, Elijah, 201–202, 373–374, 416, 422, 428
    ethnomethodology and, 437–438
Anomalous findings, 53
Anonymity, 304, 403–404, 455–457
Anspach, Renee, 276, 419
Anthony, James, 110
Antley, Angus, 84
Applied research, 20
Applied research reports, 599–605
    California Welcome Baby Program, 600–605, 602–6032e
Appropriate boundaries, 405–406
Archimedes, 222
"Are Suicidal Behaviors Contagious in Adolescence? Using Longitudinal Data to Examine Suicide Suggestion," 624–638
Arnett, Jeffrey, 472–473, 479
ARPANET, 538
Arrests for domestic violence, 42–44, 49e, 52
Art, qualitative analysis as, 416–418, 417e
Articles, research. *See* Literature, research
Asch, Solomon, 223
Aseltine, Robert, 607
Assignment variable, 236
Assisted Living Medicaid Waiver Program, 498
Association, 205–206
    describing, 341–342, 342e
    evaluating, 343–345, 343–345e
    graphing, 340–341, 340–341e

Asynchronous online interviews, 395–396
“Attitude Toward Blacks” (ATB) scale, 225, 226e, 227e
Attractions of survey research, 255–257
Attrition, differential, 240
Authenticity, 58, 61
Authority, uncritical agreement with, 10
Availability (convenience) sampling, 169–170
Avoiding harm of research participants, 76–79
Aykanian, Amanda, 480–481

Bacon, Grant A., 275
Bail, Christopher, 561
Balanced response choices, 132, 133e
Balance point, mean as, 329e, 330
Bar charts, 318, 319e, 597e
Bargh, John, 224
Barrera, Davide, 246
Base number N, 322
Basic science, 20
Battered women’s help seeking, 54–56
Baumann, Robert, 529
Baumeister, Roy, 208
Baumrind, Diana, 73–74, 77
Baxter, Caity, 495
Becker, Howard, 422–423, 440
Beecher, Henry K., 70
Before-and-after designs, 231, 235–236, 236e, 237e
“Behavioral Study of Obedience,” 73
Behavior coding, 274
Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments, 75
Bell, Frances, 16
Bellaby, Paul, 16
Bellah, Robert, 391, 393–394
Belmont Report, 71
Belousov, Konstantin, 379
Bender, Kimberly, 148, 169
Bendix, Reinhard, 553, 567
Beneficence, 71
Benwell, Bethan, 433
Berberovic, Dzenan, 375
Berdahl, Jennifer, 476

Berger, Joseph, 223
Berk, Richard A., 30, 34, 50
  benefits *vs.* risks of research by, 86–87
  on causal mechanisms, 210
  on contextual effect, 211
  deductive research, 51
  on deterrence, 245
  external validity and, 59
  field experiment, 230, 242
  hypothesis, 52
  internal validity and, 61
  Inter-university Consortium for Political and Social Research (ICPSR) and, 530
  literature review, 37, 39, 40
  mixed methods used by, 469
  selective distribution of benefits and, 247–248
  simple or complex outcomes and, 510
Berkman, Lisa F., 277
Berning, Amy, 68
Berry, Brent, 442
Besharov, Douglas, 107
Beta coefficients, 353
Bias
  coverage, 297–298
  minimizing the risk of, in survey questions, 262–263
  selection, 239–240
  self-enhancement, 204
  social desirability, 265–267
Bibliographic databases, 37
Big data, 538–543
  background on, 538–541, 540e
  ethical issues in, 544–547, 547e
  examples of research using, 541–543, 541–543e
Bimodal distributions, 328
Binge drinking, 103–105, 110–112, 116–117, 130–131
  See also Alcohol use disorder (AUD)
Bipolar response options, 265, 266e
Birckhead, Jim, 402
Birth cohorts, 199
Bivariate distributions. *See* Cross-tabulation
Black, Donald, 110
Black box evaluation, 505–506

Blair, J. E., 303
Blogging
    illogical reasoning and, 9
    overgeneralizing and, 8–9
Boas, Franz, 366
Boase, Jeffrey, 16
Bohn, Angela, 543
Bond, Robert, 546–547
Boolean search logic, 38
Boot camp programs, 510
Booth, Brenda, 163–164
Boston McKinney Project, 476, 493–494, 494e
Bouffard, Jeffrey, 297
Boundaries, appropriate, 405–406
Bourgois, Philippe, 169
Brashears, Matthew, 151
Brazilian American Cultural Institute, 114
Brenner, Katie, 232
Brewer, John, 467
British Journal of Criminology, 42–43
Broken windows thesis, 209
Brooks, Clem, 561
Brown 2013, 473–475, 479
Brown, Alfred D., 432
Brown, Ryan, 468
Buchta, Christian, 543
Bureau of Labor Statistics (BLS), 527–528
Burrows, Lara, 224
Bush, George W., 495
Bushman, Brad, 208, 230
Butler, Dore, 120
Butler, Jenny, 429
Buzawa, Eve, 56

Cain, Carole, 202, 437
Calabrese, Thomas, 611
California Welcome Baby Program, 600–605, 602–603e
Calvey, David, 403
Campbell, Alec, 30
Campbell, Colin, 312, 345
Campbell, Donald, 467
Carlisle Indian School, 442, 443e

Case oriented historical research, 554
Case-oriented understanding, 431
Case studies, 372–373
    designs, comparative, 570–571, 570–571e
    evaluating a public health program, 93–96
    likelihood of voting, 313–314, 314–315e
    patient-provider race concordance and minority health outcomes, 355–356, 356–357e
Casey, Mary Ann, 506
Caulkins, Jonathan, 513
Causality
    association and, 205–206
    context and, 210–211
    criteria and cautions for, 203–211
    idiographic, 189, 201–203
    mechanism, 209–210
    nomothetic, 189, 200–201, 201e, 203–211
    nonspuriousness and, 206–209, 207e
    in quasi-experiments, 237
    randomization and, 208, 208e
    statistical control and, 208–209, 209e
    time order and, 206
    in true experiments, 230–231
Causal network model, 425, 426e
Causal validity, 58, 60–61, 238–242
Cavanaugh, Michael, 297
Ceglowski, Deborah, 389–390
Cellphones/smartphones
    percentage of households with, 257, 257e
    surveys via, 290
    teen use of, 110
Census Bureau, U. S., 112–114, 113e, 122–123, 130, 154–156, 546
    Current Population Survey (CPS), 525
    data collected by, 526–527, 526e
    translations of surveys by, 281
Center for Epidemiologic Studies (CES-D), 137, 137e, 269, 270e
Center for Epidemiologic Studies Depression Index (CES-D), 269
Center for Mental Health Services, 120
Center for the Study of Lives, 534
Centers for Disease Control and Prevention (CDC), 93, 543
Central Intelligence Agency (CIA), 524
Central tendency, 316, 327–331, 329e, 331e

  distributions differing in variability but not, 331–332, 332e
Certificate of Confidentiality, 85
Ceteris paribus condition, 200
Chandrasekaran, Sruthi, 223
Change, resistance to, 9–10, 18
Chart Builder, 594–596e, 594–597
Chen, Mark, 224
Child and Young Adult Supplement to the National Longitudinal Survey of Youth, 254–255
Child maltreatment, 592
Chi-square, 344, 344e
Chokshi, Niraj, 232
Chronicle of Higher Education, The, 88
Civil rights movement, 112
Clark, Andrew, 444–445
Clinton, Bill, 71, 268
Clinton, Hillary, 159
Closed-ended (fixed-choice) questions, 118, 471
Clusters, 165
Cluster sampling, 165–166, 166e
Cocaine usage, graphs of, 320, 321e
Code of Ethics, ASA, 72, 246, 612–613
Code of the Street, 373–374
Coding, 316
  behavior, 274
  content, 122
  design, for variables to be measured, 579–583, 580–581e
  HyperRESEARCH, 451, 452e
  NVivo, 451–452, 453e, 454e
  stakeholders' stakes, 422, 424e
Cognitive interviews, 274
Cohen, Bernard, 223
Cohen, S., 111
Cohen, Susan, 234–235
Cohort studies, 199–200
Coin toss, 156
Coleman, James, 199–200, 467
Collaborative Institutional Training Initiative (CITI), 88
Collective efficacy, 190
Colombo, Graciela, 482
Colson, Paul, 151
Combined and compressed distributions, 325–326, 325–326e

Combined frequency displays, 325–326, 325–326e
Combined measurement options, 123–125
Combined qualitative methods, 439–440
Combined questions in indexes, 268–272
Common Rule, 72, 88
Communicating with the public, 610–611
Community-based participatory research (CBPR), 449, 450e
Community Mental Health Act Amendments, 490
Comparative methods, 560–573
  case study designs, 570–571, 570–571e
  cautions for, 571–573, 572e
  cross-sectional research, 561–565, 562–566e
  historical research, 567–569, 569e
Comparison groups, 224–225
  sources of internal invalidity reduced by, 238–239
Comparison of levels of measurement, 130–131
Compensatory rivalry, 241
Completeness, 173
Complete observers, 377
Complete participants, 377
Complex mixed-methods designs, 473–476, 474–475e
Compressed frequency displays, 325–326, 325–326e
Computer-assisted personal interviewing (CAPI), 294
Computer-assisted qualitative data analysis, 451–454, 451–454e
Computer-assisted telephone interviews (CATI), 291
Computer-mediated communication (CMC), 2–4, 13
Concepts, 104–109
  abstract and concrete, 110–115
  indicators from, 109–115, 117e
  from observations to, 115–116
  organization, categorization, and condensation in qualitative data analysis, 422–423
  overlapping dimensions of, 269, 269e
Conceptualization, 105
  in practice, 105–109
Concrete and abstract concepts, 110–115
Concurrent validity, 134
Confidence intervals, 178–179, 179e
Confidentiality and privacy, 85, 304, 403–404, 455–457
Conflicts of interest, 75
Conflict theories, 34–35
Confusing phrasing in survey questions, 260–262, 263e

Conjunctural historical research, 554
Consistent focus in surveys, 277
Constants, 112
Constructivist philosophy, 370–371
Construct validity, 134–135
Consumer Price Index (CPI), 130
Contact summary form, 421e, 422
Contamination, 241
Content, coding, 122
Content analysis, 574–575, 576–577e
    design coding, 579–583, 580–581e
    developing appropriate statistical analyses for, 583–584, 585e
    ethical issues in, 584–585
    identifying a population of documents or other textual sources for, 577–578, 578e
    selecting a sample of units for, 579
Content validity, 133
Context, 210–211
    effects, 278
    orientation to social, 369
Contingency tables, 337–339, 337–340e
Contingent questions, 262
Continuous measures, 129–130
Control groups, 205
    before-and-after designs, 231, 235–236, 236e, 237e
    ex post facto, 231, 234–235, 234e
    nonequivalent designs, 231–234
    quasi-experiments, 231–237
Convenience sampling, 169–170
Convergent validity, 135
Conversation analysis, 432–433, 434e
Cooper, Kathleen, 13
Corbin, Juliette, 428
Core Alcohol and Drug Survey, 118, 130–131
Correlational analysis, 351–353
Correlation coefficient, 352
Correll, Joshua, 120
Corroboration and legitimization of conclusions, 425–427
Cortina, Lilia, 476
Cost-benefit analysis, 503, 504e
Cost-effectiveness analysis, 503
Council of European Social Science Data Archives (CESSDA), 533

Counterfactuals, 200, 201e
"Courtesy stigma," 406
Coverage bias, 297–298
Cover letters, 283–284, 284e
Covert observers, 377, 378
Covert participation, 379–381
Cowan, Gloria, 582–583
Crabtree, Benjamin, 368, 417
Creative Interviewing, 391
Cress, Daniel, 438–439
Crime
    association and, 205–206
    broken windows thesis, 209
    committed by illegal immigrants, 187
    community treatment and diversion programs and, 188
    effect of informal social control on, 193e
    effect of social disorder and collective efficacy on rates of, 194e
    functionalism and, 35
    graphing association of, 341, 341e
    intoxication and, 204, 204e, 205–206, 205e, 211
    neighborhood disorder indicators, 446–448, 447e
    nonexperimental designs on, 206, 207e
    rates of homicides, 189e
    recidivism, 510
    self-control and, 197–198, 198e
    solitary confinement for, 432
    specific deterrence theory and, 34
    structural disadvantage and, 210, 210e
    Transitional Aid Research Project (TARP) and, 496, 498
    units of observation and, 190
Crime Control and Safe Streets Act, 515
Criterion validity, 133–134, 135
Cronbach's alpha, 137, 269
Cross-population generalizability, 59, 60e
    contingency tables, 337–339, 337–340e
Cross-sectional research designs, 192–195
    comparative, 561–565, 562–566e
    repeated, 196–197
Cross-tabulation, 337–350
    controlling for a third variable, 345–350, 346–350e
    describing association, 341–342, 342e
    evaluating association, 343–345, 343–345e

  - graphing associations and, 340–341, 340–341e
- *Current Population Survey (CPS)*, 525
- Curtis, Kristin M., 57
- Curvilinear patterns, 342
- Cyberethnography, 375
- Czopp, Alexander, 224–225, 226, 230–231
  - deception used by, 246
  - differential attrition and, 240
  - placebo effect and, 242

- Dabney, Dean, 402
- D'Amico, Elizabeth J., 499
- Dannefer, Dale, 464, 471
- Data
  - big, 538–543
  - Bureau of Labor Statistics (BLS), 527–528
  - census, 526–527, 526e
  - cleaning of, 316
  - collecting unobtrusive measures, 120–121
  - collection of qualitative, 367
  - empirical, 49–50
  - grouped, 323–325, 324e
  - Henry A. Murray Research Archive, 525, 533
  - Integrated Public Use Microdata Series (IPUMS), 527
  - international sources of, 533
  - Inter-university Consortium for Political and Social Research (ICPSR), 529–533, 530–532e, 539e, 544–545
  - links between theory and, 48e
  - other government sources of, 528
  - other sources of, 528–529
  - ownership of, 457
  - replications, 52–53
  - secondary (*See* Secondary data analysis)
  - sources of qualitative, 534
  - ungrouped, 322–323e
  - using available, 121–122
- Data analysis, 312–313
  - cross-tabulation, 337–350
  - displaying univariate distribution, 316–326
  - displaying univariate distributions, 316–326
  - ethical, 335–336, 336e, 357–358
  - frequency distributions, 322–326, 322–326e

  graphs, 318–321, 319–321e
  meta-analysis, 353–356, 356–357e
  preparing for, 315–316
  qualitative (*See* Qualitative data analysis)
  regression analysis, 351–353
  summarizing univariate distributions, 326–335
Data entry, 315–316, 317e
Data Preservation Alliance for the Social Sciences, 525
Data visualization, 596–597, 596e
Davis, Robert, 44–45
Debriefing, 82, 247
Deception, 83–84, 246–247
Decisions in research, 91–93e
Decker, Scott, 187, 453
Deductive research, 50–51
Deegan, Allison, 396
Democracy paradox, 552
Democratization in Latin America, 568, 569e
Demographic analysis, 573–574, 574e
Demographic bookkeeping equation, 573
Demography, 573
Denzin, Norman, 176, 400, 458
Department of Health and Human Services (DHHS), 72
Department of Housing and Urban Development, 524
Dependent variables, 51
Depression, 269, 270e
Descriptive research, 14–15, 56
Descriptive statistics, 313
  See also Association
Design coding procedures, 579–583, 580–581e
Desmond, Matthew, 188, 206, 208
Deterrence theory, 52–53
Detroit Area Studies, 263, 528–529
Dewey, Thomas E., 158
Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 105–106
Diamond, Timothy, 381–382, 384, 387
Dichotomies, 130
Dickson-Swift, Virginia, 406
DiClemente, Carlo, 270–271
Differences-in-differences design, 231
Differential attrition, 240
Digital ethnography, 375–376

Dill, Karen, 575, 579, 582
Dillman, 282–285
Direction of association, 51
Direct observations, 120
Discrete measures, 128
Discriminant validity, 135
Disease model, 106
Displaying research, 594–596e, 594–597
Disproportionate stratified sampling, 163–164
Distributions, sampling, 176–182, 177e, 179e
  sample size and, 180–181e, 180–182
Diversity of population, 152–154
Divorce rates, 235–236, 237e
Documentation, 420–422, 421e
Documents for content analysis, identifying populations of, 577–578
Domestic violence
  battered women's help seeking after, 54–56
  causal mechanisms, 210
  contextual effect and, 211
  formal and informal deterrents to, 40–42
  incident, offender, and outcome variables by victim-offender relationship, 56e
  inductive approach to explaining, 54
  Inter-university Consortium for Political and Social Research (ICPSR) data on, 530–533
  mass killings and, 50
  Minneapolis Domestic Violence Experiment, 28–29, 31–33, 40, 44
  Offender Focused Domestic Violence Initiative (OFDVI) and, 28
  percentage of suspects with subsequent assault by employment and arrest status, 42e
  prevalence of, 28, 29e
  research circle and, 51–53
  second responder programs and repeat family abuse incidents, 44–46, 47e
  single-article reviews on, 40–42
  when does arrest matter with, 42–44
Double-barreled questions, 262
Double-blind procedures, 241
Double negatives, 261–262
Douglas, Jack, 391
Drug Abuse Resistance Education (D.A.R.E.), 487–489, 496, 498e, 511
  impact analysis, 501e
Drug Policy Research Center (DPRC), Rand, 513–514
Duckworth, Kenneth, 579

Dugan, Laura, 402
Duggan, Maeve, 14
Duncombe, Jean, 405
Duneier, Mitch, 402, 405, 457, 567–568
Durkheim, Émile, 13, 522, 553
    functionalism and, 35
    on solidarity, 114

Earls, Felton, 190
Early childhood home visiting (ECHV), 592
Early childhood programs, 598
Eberhardt, Shelly, 192
Ecological fallacy, 190–191, 191e
EconLit, 37
Economic and Social Research Council, U. K., 534
Educational achievement, 199–200
    income and, 351–352, 352e
Eells, Kenneth, 466
Effect size, 354
Efficiency analysis, 502–504
Efficiency of survey research, 256
Ego-based commitments, 9–10
Eichmann, Adolf, 73
Elaboration analysis, 345
Elder, Keith, 399
Electronic surveys, 281–282
Elements, 149, 150
Ellis, Carolyn, 403
E-mail and social ties, 10
Embedded mixed-methods designs, 471–472
Emergence, 191
Emic focus, 415
Emmel, Nick, 444–445
Empirical data, 49–50
Empirical generalizations, 52
Employment candidates, 220–222, 222e, 243–244, 244e, 247–248
    Project New Hope and, 511, 512e
Encyclopedia Britannica, 612
Engels, Friedrich, 34
Ennett, Susan, 115, 488
Enumeration units, 150
E-Parenting Program (e-PP), 592

Equal probability of selection method (EPSEM), 160
Erbring, Lutz, 19
Erikson, Kai, 372, 381
Errors
    ecological fallacy and reductionism, 190–192, 191e
    idiosyncratic, 132
    measurement, 258–259
    random sampling, 178
    in reasoning about the social world, avoiding, 6–10
    sampling, 152, 178–179, 259
    in survey research, 258–259
    systematic, 132
Espries, Geir Arild, 233
Ethics
    achievement of valid results and, 72–74
    and considering uses of research so that benefits outweigh risks, 85–87
    in data analysis, 335–336, 336e, 357–358
    deception, 83–84, 246–247
    in evaluation, 514–516, 516e
    in experimental research, 245–248
    in historical and comparative research and content analysis, 584–585
    historical background, 70–72
    honesty and openness in, 74–75
    institutional review board (IRB) and, 87–89
    Milgram's obedience experiment and, 67–70
    mixed methods and, 481–482
    in participant observation, 379–381
    politics, and research reports, 607–613
    principles in, 72–87
    protection of research participants and, 75–85
    in qualitative data analysis, 454–458, 455e
    in qualitative research, 400–407
    in secondary data analysis and Big Data, 544–547, 547e
    selective distribution of benefits and, 247–248
    in survey research, 303–304
Ethnography, 373–374
    digital, 375–376
    video, 446, 458
Ethnomethodology, 437–438
Etic focus, 415
Ettelt, Stefanie, 514
Eurobarometer Survey Series, 533

European Community Household Panel, 522
European Social Survey (ESS), 561
Evaluability assessment, 494–495
Evaluation
    of association, 343–345, 343–345e
    black box, 505–506
    formative, 498
    of generalizability, 151–152
    of measures, 131–141
    process, 495–499, 497–498e, 499e
    of social research questions, 31–33, 32e
Evaluation research, 17, 487–489
    basics of, 491–492, 491e
    black box evaluation or program theory, 505–506, 507e
    design decisions, 504–513, 507e, 512e
    efficiency analysis, 502–504
    ethics in, 514–516, 516e
    evaluability assessment, 494–495
    history of, 490
    impact analysis, 499–502, 501e
    needs assessment, 492–494
    process, 495–499, 497–498e, 499e
    public health program case study, 93–96
    quantitative and qualitative methods in, 509–510
    questions for, 491–504, 494e, 497–499e, 501–502e
    researcher or stakeholder orientation, 506–509
    simple or complex outcomes and, 510–511, 512e
Evaluation Review, 68, 490
Evans, Tyler, 170
Event-based designs, 199–200
Events research, historical, 555–557, 556e
Event-structure analysis, 555–557, 556e
Examination and display of relationships in qualitative data analysis, 423–425, 424–425e
Excessive devotion to tradition, 10
Exhaustive choices, 118–119
Expectancies of experimental staff, 241
Experience sampling method (ESM), 386
Experimental groups, 205, 224–225
Experiments/experimentation
    causality in true, 230–231
    designs and time order, 206

    ethical issues in, 245–248
    history of, 222–224
    limitations of true, 229–230
    quasi-, 231–237, 233e
    sources of internal invalidity that require attention during, 240–242
    true, 204, 224–231
    validity in, 238–245
Explanatory research, 16, 17e, 475, 475e
    deductive research, 50–51
    inductive research, 53–54
Exploratory research, 15–16, 54–56
    questions with commitment to inductive reasoning, 367–368
Ex post facto control group designs, 231, 234–235, 234e
External events and causal validity, 238–239
External validity, 59, 244–245
Extraneous variables, 208, 347–348, 347–348e
Exum, M. Lyn, 203–204, 206, 221
    on contextual effect, 211

Facebook, 2, 3, 6, 540–541, 540e
    descriptive research about, 14–15
    experiments on its users, 544, 546–547, 547e
    surveys on, 300
    synchronous interviewing on, 396
    value of, 15e
Face validity, 132–133
Factorial surveys, 243–244
Faggiani, Don, 56
Fallon, Kathleen, 552, 567
Family Justice Center (FJC) Initiative, 495
Faulty generalizations, 105
Favorable Attitudes Toward Antisocial Behavior Scale, 127–128, 128e
Fears, Darryl, 114
Feasibility of research, 31–32
Federal Bureau of Investigations (FBI), 122, 341
Federal Interagency Committee on Education (FICE), 112
Federal Policy for the Protection of Human Subjects, 72, 88
Fedyuk, Olena, 123
Feedback, 491
Fence-sitting, minimization of, 267–268
Fenno, Richard, 382, 389, 394
Fenton, Steve, 546

Ferguson, Kristin, 148, 169, 176
Fiedler, Mora, 190
Field experiments, 230, 242
Field notes, 387–389, 388e, 420–422, 421e
Field research, 367
  developing and maintaining relationships in, 383–384
  entering the field in, 381–382
  ethnographic, 373–374
  managing the personal dimensions in, 389–390
  participant observation, 376–381
  sampling people and events in, 384–386, 385–386e
  taking notes in, 387–389, 388e
Fieldwork, 376
  See also Participant observation
Filter questions, 262, 263e
Finn, Peter, 495
First 5 LA, 600
Fischer, Constance, 431
Fisher, Bonnie, 197
Fiske, Donald, 467
Fiske, Marjorie, 397–398
Fixed-sample panel designs, 197–199, 198e
Flaming, Karl, 190
Flewelling, Robert, 488
Flickr, 441
Fligstein, Neil, 522, 524
Floaters, 267–268, 267e
Flygare, Christopher, 236
Focused Interview, The, 398
Focus groups, 274, 376, 397–399, 404
Food and Drug Administration, 72
Forced-choice questions, 268
Formal and informal deterrents to domestic violence, 40–42
Formative evaluation, 498
Forrest, Jane, 299
Foundations of social research, 35–49
Fourth Generation Evaluation, 508
Fowler, James, 546–547
Frank, David John, 558
Franke, Todd, 600
Franklin, Mark, 563
Freedom of Information Act (FOIA), 584

Frequency distributions, 322–326, 322–326e
Frequency polygons, 318, 320, 320e
Freshman, Audrey, 299e
Frohmann, Lisa, 444
Fromme, Kim, 499
Fulton, Jenna, 301
Functionalism, 35
Furr-Holden, C. Debra, 68

Galileo Galilei, 222
Gallup polls, 158, 262, 268
Gamma, 343
Gangs, youth, 106–107
Garfinkel, Harold, 438
Gatekeepers, 382
Geis, Florence, 120
Gender stereotypes, 560
Generalizability, 58, 59–60
    census and, 154–156
    evaluating, 151–152
    in qualitative research, 176, 399–400
    sample, 59, 60e, 151, 242–243
    survey research, 256–257
    in true experimentation, 242–245
Generalizations, empirical, 52
Generalizing, 8–9
General Social Survey (GSS), 177–178, 256, 258, 312, 341, 524
    International Social Survey Program, 525
    likelihood of voting and, 313–314, 314–315e
    nonresponse to, 294–295, 295e
George, Tom, 88
Gerth, Matthias, 575, 578, 579, 583
Ghetto: The Invention of a Place, The History of an Idea, 567
Gill, Hannah, 368, 389, 406
Gilligan, Carol, 436
Ginsberg, Jeremy, 543
Glueck, Eleanor, 210
Glueck, Sheldon, 210
Goal-free evaluation, 507
Gobo, Giampietro, 176, 400
Goffman, Erving, 380, 406, 440
Golder, Scott, 543

Goldstein, Adam, 522, 524
Goode, Erich, 405–406
Google Flu Trends, 543
Google Scholar, 48–49
Goulet, Lauren Sessions, 3
Governmental Accounting Standards Board (GASB), 490, 503
Grand tour questions, 392
Graphs, 318–321, 319–321e
  of association, 340–341, 340–341e
Great Depression, the, 158, 490, 557
Griffin, Larry, 554, 555–557, 556e
Grossman, Susan, 151
Grounded theory, 428–430, 429e, 430e
Group-administered surveys, 286
Grouped data, 323–325, 324e
Group randomization, 512–513
Group units of analysis, 189–190
Guba, Egon, 508
Gubrium, Jaber, 437
Guesstimates, 122
Gupta, Neeti, 5

Habits of the Heart, 391–392, 393–394
Hacker, Karen, 450
Hamilton, Edwin, 40–42, 54
Hampton, Keith N., 3, 5, 17
  applied research by, 20
  qualitative methods used by, 19
  surveys used by, 19
Hampton and gupta qualitative methods used by, 19
Handbook of Research Design and Social Measurement, 273
Haphazard sampling, 170
Hard, 611
Harris, David, 114, 555–557, 556e
Harris, Louis, 196
Harris, Lovell, 188
Harris Poll, 298
Harvard-MIT Data Center, 525
Harvard Project on Effective Interventions, 509
Harvey, Daina, 365
Hasan, Abdul, 147, 148, 149
Hastings, Orestes P., 522, 524

Hauser, Marc, 75
Hawkins, Donald, 137
Hawkins, Richard, 40
Hawthorne effect, 242
Head Start, 233–234, 390
Health and Retirement Study (HRS), 528
Health Insurance Portability and Accountability Act (HIPAA), 85, 89
Health Research Extension Act of 1985, 515
Heckman, James, 502
Henry A. Murray Research Archive, 525, 533
Herbert, Jennifer A., 202
Herek, Gregory, 91
Hermeneutic circle, 370–371, 371e
High Point Police Department (HPPD), 28
Hippocrates, 70
Hironaka, Ann, 558
Hirschel, David, 56, 61
Histograms, 318, 319e
Historical events research, 555–557, 556e
Historical process research, 557–558, 558e, 559e
Historical research methods, 551–553
    cautions for, 558–560
    comparative, 567–569, 569e
    ethical issues in, 584–585
    events research, 555–557, 556e
    overview of, 553
    process research, 557–558, 558e, 559e
    social science methods, 553–560, 555–556e, 558–559e
Historical social science methods, 553–560, 555–556e, 558–559e
History effect, 238–239
Hoffer, Thomas, 199–200
Holistic explanations, 201
Holistic historical research, 554
Hollinger, Richard, 402
Holstein, James, 437
Homelessness, Housing, and Mental Illness, 456, 477, 483
Homeless persons, 147–149
    best practices for sampling, 153
    Boston McKinney Project and, 476, 493–494, 494e
    definition of population, 151
    mean total time homeless, 479, 480e
    substance use among, 479, 479e

Homicide rates, 189e
Honesty and openness, 74–75
Hornik, Kurt, 543
Horowitz, Jonathan, 312, 345
Horrigan, John B., 16
Houtzager, Peter, 561
Howard W. Odun Institute for Research in the Social Sciences, 525
Howell, James, 107
How to Lie With Statistics, 335
Hoyle, Carolyn, 42–43
Huberman, A. Michael, 422, 423, 455
Huff, Darrell, 335
Human Population Laboratory, 277
Human Relations Area Files (HRAF), 534, 565, 566e
Human Rights Clinic, 432
Human subjectivity, 369–370
Humphreys, Laud, 70, 82–83, 380
Hunt, Dana, 495
Hunter, Albert, 467
Hurricane Floyd, 440
Hurricane Harvey, 364–365, 380
Hurricane Katrina, 365–370
    case study, 372
    focus groups, 399
    researcher safety and, 406–407
HyperRESEARCH, 451, 452e
Hypotheses, 51, 52e

Identical units, 154
Identification of social research questions, 30–31
Identity disclosure, 304, 403–404, 455–457
Idiographic explanations, 189, 201–203, 370
Idiosyncratic errors, 132
Idiosyncratic variation, 268–269
Iglesias, Gabriela, 482
Illogical reasoning, 9, 18
Immigrants, 561–562, 562–563e
Impact analysis, 499–502, 501e
Inaccurate observation, 7–8
Incentives, sampling, 173–174
Income
    education, and voting, 347–350, 348–350e

    growing inequality in, 523–524e
    by highest year of school completed, 351–352, 352e
    interpersonal trust and, 344–345, 344–345e
    voting and, 338–340, 339–340e, 346, 346e
    voting by race and, 350e
Independent variables, 51
Index, 128
Indexes, combined questions in, 268–272
Index of Status Characteristics (I.S.C.), 466
Indicators
    from concepts, 109–115, 117e
    multiple, reliability of, 136–138
Indirect communication, modern, 3
Individual matching, 233–234
Individual randomization, 512–513
Individual units of analysis, 189–190
Inductive research, 53–54
Inferential statistics, 178
Informal social control, 193e
Informants, key, 383
Informed consent, 79–83
Innstrand, Siw Tone, 233
In-person interviews, 293–295, 295e
Inputs, 491
Institute for Survey Research, 96
Institutional review board (IRB), 87–89, 545
    case study, 93–96
Instrumentation and causal validity, 239
Integrated literature reviews, 42–44
Integrated mixed-methods design, 470–471
Integrated Public Use Microdata Series (IPUMS), 527
Integrative approaches, 508
Integrity and quality, research, 457
Intelligence Quotient (IQ), 129
Intensive (in-depth) interviewing, 376, 390–395
Intent-to-treat analysis, 240–241
Interaction of testing and treatment, 245
Interactive voice response (IVR), 292, 292e
Intercensal percent change, 573
Intercoder reliability, 138
Interitem reliability, 136–137
Internal validity, 58, 60–61, 238–242

International data sources, 533
International Federation of Data Organizations for Social Science (IFDO), 533
International Social Survey Program (ISSP), 525, 561
International Typographical Union, 469
International Visual Sociology Association (IVSA), 442
Internet, the, 2–4
  activity in one minute, 540e
  communication options on, 454, 455e
  community life and access to high-speed, 17
  effect on social relations, 16
  interviewing on, 395–397
  overgeneralization based on, 8–9
  percentage of individuals not using, by personal characteristics, 5, 5e
  percentage of U.S. adults who use, 3–4, 4e
  quantity of data generated on, 540–541, 540e
  resources on heart conditions on, 16
  searching the, 46–49
  surveys via, 296–300, 296e, 299e
Interobserver reliability, 138, 139e
Interpretive comments, 418
Interpretive questions, 275–276
Interpretivism, 20
Interquartile range, 333
Intersubjective agreement, 20
Inter-university Consortium for Political and Social Research (ICPSR), 525, 529–533, 530–532e, 539e, 544–545
Interval, sampling, 162
Interval level of measurement, 126e, 128–129, 131e
Interval-ratio level of measurement, 131
Intervening variables, 346–347, 346e
Intervention and advocacy, 457
Interviewing
  in-person, 293–295, 295e
  intensive, 376, 390–395
  online, 395–397
Interview schedules, 273, 279e
Intimate partner violence (IPV), 28
  defined, 58
  See also Domestic violence
Intoxication. *See* Alcohol use disorder (AUD) and intoxication
Intraobserver reliability, 136
Intrarater reliability, 136

Irvine, Leslie, 403
Izmit earthquake, 369, 371–372

Jalbert, Sarah Kuck, 236
James, Erica, 406
Jang, Bohyun Joy, 103, 116, 149
Jessop, Julie, 405
Job Training Partnership Act (JTPA), 500
John Henry effect, 241
Jones, Jim, 73
Jottings, 387
Journal articles, 598–599, 600e
Journal of Abnormal and Social Psychology, 73
Journal of Research in Crime and Delinquency, 488
Judd, Charles, 120
Jude, Frank, 188
Justice, 71

Kale-Lostuvali, Elif, 369, 371–372, 376, 440
Kane, Michael, 236
Kato, Yuki, 365
Kemmis, Stephen, 449
Kendall, Patricia, 397–398
Kessler, Ronald, 607
Key informants, 383
Kilgore, Sally, 199
Kiliç, Ayse, 39
King, Valarie, 137
Kippen, Sandra, 406
Kirk, David, 188, 206, 208
Klap, Ruth, 30
Klinenberg, Eric, 384
Kolbert, Elizabeth, 398
Kollock, Peter, 16
Kozinets, Robert, 375
Kuhn, Thomas S., 35

Labaw, Patricia J., 275
Labeled unipolar response options, 265, 266e
Labeling theory, 31, 34
Landers, Ann, 7
Landline phones, 257, 257e
Landon, Alfred M., 158

Laub, John, 197, 198
    on causal mechanisms, 210
    on contextual effect, 211
Law enforcement
    masculinity and, 426–427
    race and, 188, 192
Layte, Richard, 522
Leakey, Tricia, 515–516
LeBlanc, Jessia, 18
Ledford, Gerald, 234–235
Legal cynicism, 188
Lettiere, Mark, 169
Levels of measurement, 125–131, 126e
    comparison of, 130–131
    dichotomies, 130
    interval, 126e, 128–129, 131e
    nominal, 126–127, 126e, 131e
    ordinal, 126e, 127–128, 128e, 131e
    properties of, 131e
    ratio, 126e, 129–130, 131e
Levinson, Martin, 401, 402, 455
Levitt, Heidi, 429
Lewis, Kevin, 19
Liamputtong, Pranee, 406
Library of Congress, 525
Lieven, Filip, 243–244
Likert items, 265
Lincoln, Yvonna, 508
Lindsay, Sally, 16, 122
Ling, Rich, 5, 18, 151
    quantitative methods used by, 19
Link, Bruce G., 161
Lipset, Seymour Martin, 467, 469
Literal focus, 418
Literary Digest, 158
Literature, research
    how to read, 619–638
    integrated reviews, 42–44
    questions to ask about articles in, 617–618
    reviewing, 39–46
    searches, 36–39
    single-article reviews, 40–42

- systematic reviews, 44–46
- on the web, 46–49

Livingston, Jennifer, 465
Longitudinal research designs, 192–193, 195–200, 196e
Luxardo, Natalia, 482
Lynching, 555–557, 556e

Maccio, Elaine, 148, 169
Macy, Michael, 543
Magazine articles, 36–37
Mahoney, James, 567
Mailed, self-administered surveys, 282–286, 284e
Mair, Patrick, 543
Majchrzak, Ann, 513–514
Malinowski, Bronislaw, 366
Manza, Jeff, 561
Marginal distributions, 338
Margolis, Eric, 442
Mark, Aimee, 224
Markus, M. Lynne, 513–514
Marshall, Gary, 83
Martins, Silvia, 110
Marx, Karl, 34, 522
Massachusetts Department of Public Health, 93
Mass killings and domestic violence, 50
Matching, 227, 229e
- aggregate, 232–233
- individual, 233–234

Mateen, Omar, 50
Matrix, 423, 424e
Matrix questions, 279
Maturation and causal validity, 239
Mays, Nicholas, 514
McCain, John, 158
McCarter, Susan, 470, 477
McClelland, Charles, 529
McLellan, A. Thomas, 134–135
McPherson, Miller, 18, 151
McTaggart, Robin, 449
Mean, 328–329, 329e
- median or, 329–330

Measures/measurement, 116–125

  of association, 343
  central tendency, 316, 327–331, 329e, 331e
  coding content, 122
  constructing questions for, 118–119
  continuous, 129–130
  defined, 116
  design coding procedures for, 579–583, 580–581e
  dichotomies, 130
  error, 258–259
  evaluating, 131–141
  levels of, 125–131, 126e
  making observations for, 119–120
  options, combined, 123–125
  photographic, 122–123, 123e
  reliability, 135–138
  unobtrusive, 120–121
  using available data, 121–122
  validity, 58–59, 132–135, 138–141
  of variation, 331–335
Mecca, Laurel Person, 441
Mechanical Turk, 243
Mechanism, 209–210
Median, 328, 329–330
Mediators, 209
MEDLINE, 37, 38, 40
Medway, Rebecca, 301
Meeker, Marchia, 466
Mejia, Camilia, 125
Merton, Robert K., 397–398
Meta-analysis, 353–356, 356–357e
Method of agreement, 570–571, 570e, 572e
Method of difference, 570–571, 571e, 572e
Michigan Alcoholism Screening Test (MAST), 133
Miles, Matthew, 422, 423, 455
Milgram's obedience experiment/Stanley Milgram, 67–70, 76, 77, 149
  achievement of valid results in, 72–74
  benefits *vs.* risks of, 85–86
  deception in, 83–84, 246
  diversity of population in, 153–154
  honesty and openness of, 74–75
  informed consent in, 82
Mill, John Stuart, 570–571e

Miller 1999, 426–427
Miller, Arthur, 79
Miller, Delbert, 273
Miller, JoAnn, 86
Miller, Susan, 419–420
Miller, William, 368, 417
Mills, C. Wright, 30
Minneapolis Domestic Violence Experiment, 28–29, 31–33, 40, 44, 510
    as deductive research, 51, 52–53
    as descriptive research, 56
    See also Berk, Richard A.; Sherman, Lawrence
Mirowsky, John, 256, 273, 304
    phone survey, 287, 292
    survey questions, 262, 264, 268, 278
Mitchell, Richard, Jr., 381
Mixed methods, 464–466
    complex designs, 473–476, 474–475e
    embedded designs, 471–472
    ethics and, 481–482
    history of, 466–468
    integrated designs, 470–471
    staged designs, 472–473
    strengths and limitations of, 476–481, 477e, 479–480e
    types of, 468–476, 469e, 474–475e
Mixed-mode surveys, 300–301
Mobile phones and social ties, 5
Mode, 327–328
Moe, Angela, 54–55, 61, 469
Monitoring the Future, 116
Monotonic patterns, 342
Monteith, Margo, 224
Moore, Barrington, Jr., 553
Moore, Spencer, 440
Moral tales, 435–436
Morgan, Philip, 557
Morrill, Calvin, 413, 414–415, 435–436
Motives for social research, 13
Mueller, Anna S., 624
Multidimensional Poverty Index (MPI), 107
Multi-Method Research, 467
Multiple group before-and-after design, 235
Multiple regression analysis, 353, 353e

Multistage cluster sampling, 166, 166e
Multitrait-multimethod matrix, 467
Mutually exclusive choices, 118–119
Mykletun, Reidar, 233
My Lai massacre, Vietnam, 73

Nakonezny, Paul A., 235–236
Narrative analysis, 435–437, 436e
Narrative explanations, 554
National Archive of Criminal Justice Data, 525
National Archives and Records Administration, 525
National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 71
National Data Program for the Social Sciences, 525
National Digital Preservation Program, 525
National Geographic Society, 257, 298
National Incident-Based Reporting System (NIBRS), 122
National Institute of Child Health and Human Development (NICHD), 528
National Institute of Justice, 495
National Institute of Mental Health, 140, 476, 478
National Institute on Aging (NIA), 528
National Institute on Alcohol and Alcoholism (NIAAA), 105
National Institute on Drug Abuse, 110
National Institutes of Health (NIH), 85, 87
National Longitudinal Survey of Youth, 254–255, 269, 277
National Longitudinal Surveys (NLS), 528
National Minority SA/HIV Prevention Initiative Youth Questionnaire, 129
National Opinion Research Center (NORC), 255, 524
National Roadside Survey, 68
National Science Foundation, 525
National Survey of Families and Households, 528
National Technical Information Service (NTIS), 528
National Violence Against Women Survey, 59
Nature, 612
Nayak, Manan, 90
Nazi Germany, 70, 73, 567
Needleman, Carolyn, 464, 465–466, 469
Needs assessment, 492–494
“Negative Outlook” index, 137, 137e
Negatives, double, 261–262
Netnography, 375
Neuendorf, Kimberly, 575, 578, 579

Newbury, Darren, 444
New England Journal of Medicine, 70
New Jersey Income Maintenance Experiment, 490, 495–496
Newport, Frank, 268
Newspaper articles, 36–37
Newsweek, 320
New York Times, 546, 579
Ngrams, 541–542, 541e
Nie, Norman H., 19
Nisbett, Richard, 88
Nixon, Richard M., 262
Nobles, Matt, 297
Nolte, Ellen, 514
Nominal level of measurement, 126–127, 126e, 131e
Nomothetic causal explanations, 189, 200–201, 201e
    criteria and cautions for, 203–211
Nonequivalent control group designs, 231–234
Nonexperimental designs and time order, 206, 207e
Nonprobability sampling methods, 156, 168–174
    availability (convenience) sampling, 169–170
    purposive sampling, 172–173
    quota sampling, 170–172, 171e
    snowball sampling, 173–174
Nonrespondents, 157
Nonresponse, 259
    to interviews, 294–295, 295e
    to mailed, self-administered surveys, 283
    to phone surveys, 290, 290e, 291e
Nonspuriousness, 206–209, 207e
Nordanger, Dag, 367–368, 370, 376, 378–379, 382, 392
Normal distribution, 334–335, 335e
Normal science, 35
Note taking, 387–389, 388e, 420–422, 421e
Nuremberg War Crime Trials, 70
NVivo, 451–452, 453e, 454e

Obama, Barack, 158, 313–314
O'Brien, Dan, 542
Observation, 7–8, 8e
    complete, 377
    to concepts, 115–116
    covert, 377, 378

    direct, 120
    measurement using, 119–120
    overt, 378
    systematic, 446–448, 447–448e
    units of, 190
Observation, participant, 376
    choosing a role for, 377–381
    continuum, 377, 377e
    developing and maintaining relationships in, 383–384
    entering the field for, 381–382
    field notes and, 387–389, 388e, 420–422, 421e
    managing the personal dimensions in, 389–390
    sampling of people and events in, 384–386, 385–386e
Offender Focused Domestic Violence Initiative (OFDVI), 28
Office for Protection From Research Risks, 87
Ogburn, William Fielding, 21
Oklahoma City terrorist bombing, 235–236, 237e
Olson, Kristen, 301
Omnibus surveys, 258
Ondersma, Steven, 592
Online interviewing, 395–397
Onoye, Jane M., 298
Open-ended questions, 119, 276, 471
Openness and honesty, 74–75
Operationalization, 109
    of concept of race, 112–114
    of social network position, 114–115
Optical illusions, 7e
Oral history, 557
Ordering of questions, 277–279, 279e
Ordinal level of measurement, 126e, 127–128, 128e, 131e
Organic solidarity, 35
Organisation for Economic Co-operation and Development (OECD), 524, 561
Organizations, social research, 57
Orientations, research, 19–21
Orshansky, Mollie, 107
Outcomes, program, 491
Outliers, 333
Output, 491
Overgeneralization, 8–9, 8e
Overt observation, 378
Overt participation, 377, 378–379

Ownership of data and conclusions, 457

Pacula, Rosalie Liccardo, 513
Pagnini, Deanna, 557
Panel studies, 197–199, 198e
Panel Study of Income Dynamics, 240
Papachristos, Andrew, 188, 206, 208
Paradigms, 35
“Paradigm wars,” 468
Park, Bernadette, 120
Partially spurious relationships, 207e
Participant observation, 376
    choosing a role for, 377–381
    continuum, 377, 377e
    developing and maintaining relationships in, 383–384
    entering the field for, 381–382
    field notes and, 387–389, 388e, 420–422, 421e
    managing the personal dimensions in, 389–390
    sampling of people and events in, 384–386, 385–386e
Participant observers, 377, 378–379
Participants, protection of research, 75–85
    avoiding deception in research, except in limited circumstances, and, 83–84
    avoiding harm in, 76–79
    maintaining privacy and confidentiality in, 85
    obtaining informed consent in, 79–83
Participant well-being, 402–403
Participatory action research, 449–450, 450e, 455, 508
Part-whole question effect, 278
Passidomo, Catarina, 365
Pate, Antony, 40–42, 54
Paternoster, Raymond, 34, 44, 50
Patient-provider race concordance and minority health outcomes, 355–356, 356–357e
Patrick, Megan, 103, 149
Pattavina, April, 56
Patterson, Orlando, 113
Patton, Michael Quinn, 509
Patulny, Roger, 2, 3, 16, 19, 149
    basic science used by, 20
    quantitative methods used by, 19
Pauwels, Luc, 443–444
Pearce-Morris, Jennifer, 254–255

Pearson, Frederic, 529
Pearson, Michael, 114
Pedulla, David S., 221, 244
    field experiment, 230
Peer-reviewed articles, 36
Pepinsky, Harold, 379
Percentages, 338–339, 339e
Perry, Gina, 75, 77
Personal identification numbers (PIN), 297
Personal motivations for social research, 13–14, 14e
Pew Internet & American Life Project, 14, 16, 19
Pew Research Center, 4–6, 255, 525–526
Phillips, David, 235
Philosophical perspectives, 20
Phone surveys, 286–293, 289–292e
Photographs, 122–123, 123e, 441
Photo voice, 444, 444e
Piliavin, Irving, 84
Piliavin, Jane Allyn, 84
Placebo effect, 241–242
Plagiarism, 611–613
Planetary Society, 260
Planning, sample, 149–156
Plessy v. Ferguson, 112
Policy motivations for social research, 13
Policy research, 513–514
Politics and reporting, 607–613
Pollio, David, 148, 169
Polls, 158–159, 159e, 161e, 327
    effect of floaters on, 267–268, 267e
    push, 263
Poor measurement, 258–259
Population, 149
    assessing diversity of, 152–154
    coverage bias, 297–298
    defining, 151
    inadequate coverage of, 259
    sample components and, 149–151
    sampling of vulnerable, 153
    selecting a sample of units from, for content analysis, 579
    target, 152
Population change, 573

Population composition, 573–574, 574e
Population of documents for content analysis, identifying, 577–578
Population parameter, 178
Population pyramids, 573–574, 574e
Positivism, 20
Posters, research, 605–607, 606e, 608e
Postpositivism, 20
Posttest measures, 225–226
Poverty
    conceptualization of, 107–109
    contextual effect, 211
    motivation for studying immigrants living in, 13, 14e
    relative, 107–108
    State Children's Health Insurance Program (SCHIP) and, 500, 502e
PowerPoint, 596, 596e
Precoding, 316
Predictive validity, 134
Prejudice, 105
Preschool teachers, 598
Presidential election polls, 158–159, 159e, 161e, 327
Pretest measures, 225–226
Pretests, survey, 274
Price, Richard, 239, 241, 245, 247–248
Primary sampling units, 150
Priming effect, 224
PRISMA guidelines, 44, 46e
Privacy and confidentiality, 85, 304, 403–404, 455–457
Probability average, 327–328
Probability of selection, 156
Probability sampling methods, 156, 157–168
    cluster sampling, 165–166, 166e
    compared, 167–168, 167–168e
    simple random sampling, 160–161, 161e
    stratified random sampling, 162–165, 164e, 165e
    systematic random sampling, 162, 163e
Procedural justice theory, 34
Proceedings of the National Academy of Sciences, 192
Process consent, 402
Process evaluation, 495–499, 497–498e, 499e
Process research, historical, 557–558, 558e, 559e
Program process, 491
Program theory, 505–506, 507e

Progressive focusing, 416
Project New Hope, 511, 512e
Project SPLASH (Smoking Prevention Launch Among Students in Hawaii), 515–516
Proportionate stratified sampling, 163
Proposals, social research, 90–96
Protection of research participants, 75–85
  avoiding deception in research, except in limited circumstances, and, 83–84
  avoiding harm in, 76–79
  maintaining privacy and confidentiality in, 85
  obtaining informed consent in, 79–83
Pseudoscience, 12–13
PsycARTICLES, 37
Psychological Abstracts, 37
PsycINFO, 37, 40
Public Agenda, 509
Public Health Service, U. S., 70
Purcell, Kirsten, 3
Purposive sampling, 172–173
Push polling, 263

Qualitative comparative analysis (QCA), 438–439, 439e
Qualitative data analysis, 414–415
  abductive analysis, 430–431
  alternatives in, 428–440, 429–430e, 434e, 436e, 439e
  as an art, 416–418, 417e
  case-oriented understanding, 431
  causal network model, 425, 426e
  compared with quantitative, 418–419
  computer-assisted, 451–454, 451–454e
  conversation analysis, 432–433, 434e
  corroboration and legitimization of conclusions in, 425–427
  documentation, 420–422, 421e
  ethics in, 454–458, 455e
  ethnomethodology, 437–438
  examination and display of relationships in, 423–425, 424–425e
  features of, 415–419, 417e
  grounded theory, 428–430, 429e, 430e
  narrative analysis, 435–437, 436e
  organization, categorization, and condensation in, 422–423
  participatory action research, 449–450, 450e, 455
  qualitative comparative analysis, 438–439, 439e

- reflection on researcher's role in, 427–428
- systematic observation, 446–448, 447–448e
- techniques of, 419–428, 420–421e, 424–426e
- visual sociology, 440–446

Qualitative Data Archiving Resource Center, 534

Qualitative data sources, 534

Qualitative research, 19–20
- acceptance in, 370–371
- adaptive research design in, 371–372
- appropriate boundaries, 405–406
- basics of, 372–376
- case study, 372–373
- combining methods of, 439–440
- data collection, 367
- digital ethnography, 375–376
- ethical issues in, 400–407
- ethnography, 373–374
- evaluation research, 509–510
- exploratory research, 15–16, 54–56, 367–368
- features of, 367–372
- focus groups, 274, 376, 397–399, 404
- focus on human subjectivity, 369–370
- focus on previously unstudied processes and unanticipated phenomena, 367–368
- fundamentals of, 366–376
- generalizability in, 176, 399–400
- history of, 366–367
- idiographic explanations, 189, 201–203, 370
- intensive interviewing, 376, 390–395
- interviewing online, 395–397
- orientation to social context, 369
- participant observation, 376–390
- reporting, 607
- researcher safety in, 406–407
- sensitivity to subjective role of researcher in, 368–369, 369e
- *See also* Explanatory research

Quality
- sample, 174–175
- secondary data, 536

Qualtrics, 299–300

Quantitative analysis. *See* Data analysis

Quantitative research, 19–20

  content analysis, 122
  evaluation research, 509–510
  nomothetic causal explanations, 189, 200–201, 201e, 203–211
  reporting, 607
  See also Data analysis
Quartiles, 333
Quasi-experiments, 231–237, 233e
  causality in, 237
Quesada, James, 169
Questionnaires
  design of, 273–281
  made attractive, 280
  translation of, 280–281
Questions, social research, 30–33
  asking and recording answers to, 393–395
  constructing, 118–119
  evaluation research, 491–504, 494e, 497–499e, 501–502e
  grand tour, 392
  literature searches and, 37
  survey (*See* Survey questions)
Quota sampling, 170–172, 171e, 172e

Race
  "Attitude Toward Blacks" (ATB) scale and, 225, 226e, 227e
  law enforcement and, 188, 192
  lynching and, 555–557, 556e
  operationalizing the concept of, 112–114
  out-of-wedlock births and, 557
  patient-provider race concordance and minority health outcomes, 355–356, 356–357e
  political parties and, 398
  by region of the U.S., 340–341, 340–341e
  voting by income and, 350e
  wedding party photos and, 442, 443e
Rainie, Lee, 3, 14–15, 16, 525–526
Randall, Ann, 396–397
Random digit dialing, 160
Random errors, 132
Randomization/random assignment, 208, 208e, 226–228, 227–228e, 229e
  group, 512–513
  sources of internal validity reduced by, 239–240
Randomized comparative change design, 225

Randomized comparative posttest design, 226
Random number tables, 160
Random sampling, 156–157, 228e
    error, 178
    simple, 160–161
    stratified, 162–165, 164e, 165e, 172e
    systematic, 162, 163e
Range, 332–333
    interquartile, 333
Rapport and control in in-person interviews, 293–294
Ratio level of measurement, 126e, 129–130, 131e
Rational choice theory, 33e, 34
Raudenbush, Stephen, 190, 441, 442
    causal conclusions by, 204
    cross-sectional designs, 193–194
    nonexperimental designs, 206, 208
    statistical control and, 209
    study of association, 205–206
    systematic observation by, 446–448, 447e
Reactive effects, 378
Reasoning, 9, 18
Recording answers to questions, 393–395
Reddick, Rebecca, 235–236
Reductionism, 191–192
Reevaluating, 9–10
Refereed journals, 36
Refining of social research questions, 31
Reflexive comments, 418
Reflexivity, 368–369, 427–428
Regression and causal validity, 239
Regression-discontinuity design, 236
Regression effect, 239
Reisman, David, 13
Reiss, Albert, 120
Relative poverty, 107–108
Reliability, measurement, 135–138
    multiple indicators, 136–138
    multiple observers, 138, 139e
    multiple times, 136
    ways to improve, 138–141
Repeated cross-sectional designs, 196–197
Repeated measures panel designs, 236

Replacement sampling, 160
Replications, 52–53
Reporting
    communicating with the public in, 610–611
    ethics in, 607–613
    plagiarism and, 611–613
Reporting research, 597–607, 597e, 600e, 602–603e, 606e
    applied research reports, 599–605
    on California Welcome Baby Program, 600–605, 602–603e
    in journal articles, 598–599, 600e
    quantitative and qualitative research, 607
    in research posters, 605–607, 606e, 608e
    well-written, 613
Representative samples, 154, 155e
Request for proposals (RFP), 31
Research circle, 50, 51e
    domestic violence and, 51–53
Research designs, 189–203
    adaptive, 371–372
    comparing, 211–212, 211e
    comparison of survey, 301–303, 302e
    cross-sectional and longitudinal, 192–200, 196e
    evaluation, 504–513, 507e, 512e
    mixed methods (*See* Mixed methods)
    quantitative or qualitative causal explanations and, 200–203
    quasi-experimental, 231–237
    units of analysis, 189–192
    See also Experiments/experimentation
Researcher or stakeholder orientation in evaluation research, 506–509
Researcher safety, 406–407
Research literature. *See* Literature, research
Research participants. *See* Protection of research participants
Research posters, 605–607, 606e, 608e
Research program with mixed methods, 469
Research spiral, 54, 55e
Resistance to change, 9–10, 18
Respect for persons, 71
Respondent-driven sampling, 173–174, 174e, 175e
Responsive evaluation, 507
Results of research
    displaying, 594–596e, 594–597
    reporting, 597–607, 597e, 600e, 602–603e, 606e

  - writing, 592–594
- Reuter, Peter, 513
- Reverse outlining, 593
- Reviewing of literature research, 39–46
- Rhodes, William, 236
- Riessman, Catherine Kohler, 202, 437
- Rinehart, Jenny, 414
- Ringwalt, Christopher, 488, 496
- Risk Skills Training Program (RSTP), 499
- Robert Wood Johnson Foundation, 489, 607
- Rodgers, Joseph Lee, 235–236
- Rodriguez, Havidan, 370, 376
- Roles in participant observation, 377–381
- Roosevelt, Franklin D., 158
- Roper Center, 255, 525
- Rosenthal, Rob, 173
- Ross, Catherine, 256, 273, 280, 303, 304
  - phone survey, 287, 292
  - survey questions, 262, 264, 268, 278
- Rossi, Peter, 167, 496, 512, 610–611
- Rubin, Herbert, 172–173
- Rubin, Irene, 172–173
- Rueschemeyer, Dietrich, 568

- Sacks, Stanley, 503
- Safe and Drug-Free Schools and Communities Act, 490
- Safety, researcher, 406–407
- Sahiner, Gönül, 39
- Salkind, Neil J., 273
- Salmons, Janet, 404
- Sample components, 149–151, 150e
- Sample generalizability, 59, 60e, 151, 242–243
- Sample planning, 149–156
- Sample(s), 149
  - quality of, 174–175
  - representative, 154, 155e
  - of units selected from the population for content analysis, 579
- Sample size, 180–181e, 180–182
- Sample statistic, 178
- Sampling
  - frame, 149
  - methods (*See* Sampling methods)

    of people and events in field research, 384–386, 385–386e
    purpose of, 149
    random, 156–157, 228e
    theoretical, 385–386, 386e
    units of, 150
Sampling distributions, 176–182, 177e, 179e
    sample size and, 180–181e, 180–182
Sampling error, 152, 259
    estimating, 178–179
Sampling interval, 162
Sampling methods, 156–176
    nonprobability, 168–174
    probability, 156, 157–168
Sampson, Robert, 190, 191–192, 441, 442
    causal conclusions by, 204
    on causal mechanisms, 210
    on contextual effect, 211
    cross-sectional design, 193–194
    fixed-sample panel design, 197, 198
    nonexperimental designs, 206, 208
    statistical control and, 209
    study of association, 205–206
    systematic observation by, 446–448, 447e
Sanders, Andrew, 42–43
Sandstrom, Heather, 600
Saturation, 173
Saturation point, 392, 393e
Scales, 272
Schleyer, Titus, 299
Schofer, Evan, 558
Schofield, Janet Ward, 400
Scholarly journals, 36
School cohorts, 199
Schorr, Lisbeth, 509–510
Schreck, Christopher, 197
Schreer, George, 84
Schuler, Megan, 103, 149
Schur, Lisa, 232
Schutt, Julia, 610
Schutt, Russell, 119, 456, 471, 478
Science, 10–17
    basic, 20

  defined, 10
  motives for social research, 13
  normal, 35
  pseudoscience or, 12–13
  types of social research and, 14–17
Scientific approach, 10–12
Scientific paradigms, 35
Scientific relevance of social research, 33
Seaman, Claire, 2, 3, 16, 19, 149
  basic science used by, 20
  quantitative methods used by, 19
Seattle–Denver Income Maintenance Experiment, 490
Sechrist, Stacy, 28, 39
Secondary data analysis, 522
  challenges for, 534–538
  ethical issues in, 544–547, 547e
  sources of data for, 522–534
Secondary sampling units, 150
Second Life, 396–397
Second responder programs and repeat family abuse incidents, 44–46, 47e
Selection bias, 239–240
Selective distribution of benefits, 247–248
Selective observation, 7, 8e
Self-administered, mailed surveys, 282–286, 284e
Self-control and crime, 197–198, 198e
Self-enhancement bias, 204
Self-reports, inadequacy of, 124, 124e
Seniority cohorts, 199
Serendipitous findings, 53
Setting, survey, 282
Sexual abuse/assault
  campus, 465–466
  prevalence of, 28, 29e
  in the workplace, 476
Sexual Experiences Survey (SES), 414
Sherman, Lawrence, 34, 50
  benefits *vs.* risks of research by, 86–87
  on causal mechanisms, 210
  on contextual effect, 211
  deductive research, 51
  on deterrence, 245
  external validity and, 59

  field experiment, 230, 242
  hypothesis, 52
  internal validity and, 61
  Inter-university Consortium for Political and Social Research (ICPSR) and, 530
  literature reviews, 37, 39, 40
  mixed methods used by, 469
  selective distribution of benefits and, 247–248
  simple or complex outcomes and, 510
Sidewalk, 457
Siegert, Gabriele, 575, 578, 579, 583
Sim, Jeremiah, 114
Simple or complex outcomes, 510–511, 512e
Simple random sampling, 160–161, 161e
Simpson, Brett, 246
Sin, Chih Hoong, 482
Single-article reviews, 40–42
Size, sample, 180–181e, 180–182
Skewness, 316
Skip patterns, 262, 263e
Skocpol, Theda, 568, 570
Slater, Mel, 84
Sloboda, Zili, 489, 511, 515
Smith, Aaron, 14
Smith, Marc, 16
Smith, Saundra, 84
Smith, Simon, 16
Smith-Lovin, Lynn, 151
Smyth, Jolene, 301
Snow, David, 438–439
Snowball sampling, 173–174
Sobell, Linda, 136
Social Care Institute for Excellence, 490–491
Social Class in America, 466
Social control, 109e, 110
  effect of informal, 193e
Social desirability bias, 265–267
Social health, 104–105
Social importance of research, 32
Social isolation, 5
Socially desirable behavior, 124, 124e
Social movement organizations (SMO), homeless, 438–439, 439e

Social Networking Sites and Facebook Survey, 3–6
Social networking/social media, 2–4
    and changes in size of social networks, 6e
    interviewing through, 395–397
    position, operationalizing of, 114–115
    quantity of data generated on, 540–541, 540e
    surveys via, 300
Social Origins of Democracy and Dictatorship, 553
Social research
    alternative research orientations for, 19–21
    foundations, 35–49
    motivations for, 13
    questions, 30–33
    strategies, 49–56
    strengths and limitations of, 18–19
    types of, 14–17
Social research organizations, 57
Social research proposals, 90–96
Social research standards, 57–61
Social Science Citation Index (SSCI), 37
Social sciences
    approaches in evaluation research, 507
    comparative methods, 560–573
    definition of, 12
    historical methods, 553–560, 555–556e, 558–559e
    survey research in, 255–259
Social Security Administration, 533
Social status, 111
Social stratification, 422
Social support, 111
Social theories, 30–31, 33–35
    links between data and, 48e
    and predictions about effect of arrest for domestic assault, 49e
    social status in, 111
Social ties
    e-mail and, 10
    explanatory research on Internet effect on, 16, 17e
    mobile phones and, 5
    mortality and, 277
Social world, the
    avoiding errors in reasoning about, 6–10
    learning about, 4–6

SocINDEX, 37–38, 40, 533
Socioeconomic status, 116
Sociological Abstracts, 37, 40, 41e
Sociology, visual, 440–446
Socius, 524
Solitary confinement, 432
Solomon four-group design, 228, 245e
Sosin, Michael, 151
Specification, 348–350, 349–350e
Specific deterrence theory, 34
Split-ballot decision, 258
Split-half reliability, 137–138
Spurious relationships, 207–208, 207e
    use of statistical control to reduce, 209e
St. Jean, Peter, 448, 448e
St. Pierre, Robert, 512
Stack, Angela, 208, 230
Staged mixed-methods designs, 472–473
Stake, Robert, 372, 422
Stakeholders, 491
    approaches in evaluation research, 507–508
Stald, Gitte, 5, 18, 151
    quantitative methods used by, 19
Standard deviation, 333–335, 335e
Standardization of population numbers, 573
Standards, social research, 57–61
Stanford Research Institute, 538
State Children's Health Insurance Program (SCHIP), 500, 502e
States and Social Revolutions, 571
Statistical control, 208–209, 209e
Statistical Package for the Social Sciences (SPSS), 316
Statistical power analysis, 181
Statistical significance, 344
Statistics, 313–314, 314–315e
    cross-tabulating variables, 337–350
    displaying univariate distributions, 316–326
    ethical data analysis, 335–336, 336e
    inferential, 178
    measures of central tendency, 316, 327–331, 329e, 331e
    meta-analysis, 353–356, 356–357e
    regression analysis, 351–353
    summarizing univariate distributions, 326–335

    See also Association
Stephen, Jason, 611
Steward, Eric, 197
Stokoe, Elizabeth, 433, 441, 442
Strategies, social research, 49–56
    See also Explanatory research; Exploratory research
Stratified random sampling, 162–165, 164e, 165e, 172e
Strauss, Anselm, 428
Strauss, Jaine, 575, 578, 579, 583
Street Corner Society, 382
Street-level bureaucrats, 535–536
Strengths and limitations of social research, 18–19
Stress among individuals with intellectual disabilities, 233, 233e
Structural disadvantage and crime, 210, 210e
Structure of Scientific Revolutions, The, 35
Study of American Families (SAF), 312
Subject descriptors, 38
Subject fatigue, 198–199
Subjective role of researchers in qualitative research, 368–369, 369e
Subjective well-being, 108–109
Subject well-being, 402–403
Substance use
    among young adults, 103–104
    applied reports, 605, 606e
    cocaine usage graphs, 320, 321e
    conceptualization in practice, 105–106
    Drug Abuse Resistance Education (D.A.R.E.), 487–489, 496, 498e, 511
    homelessness and, 479, 479e
    self-report and observer ratings of, 477–478, 477e
Subtables, 346
Suicides
    out-of-wedlock births and, 557
    TV soap operas and, 235, 236e
Survey Monkey, 299–300, 299e
Survey organization, 281–303
    cost, 282
    cover letters, 283–284, 284e
    group-administered surveys, 286
    in-person interviews, 293–295, 295e
    mailed, self-administered surveys, 282–286, 284e
    manner of administration, 281–282
    mixed-mode surveys, 300–301

  questionnaire structure, 282
  setting, 282
  telephone surveys, 286–293, 289–292e
  web surveys, 296–300, 296e, 299e
Survey pretests, 274
Survey questions, 259–268
  avoiding confusing phrasing in, 260–262, 263e
  avoiding making either disagreement or agreement disagreeable in, 265–267
  combined in indexes, 268–272
  interpretive, 275–276
  maximizing the utility of response categories in, 264–265
  minimizing fence-sitting and floating in, 267–268, 267e
  minimizing the risk of bias in, 262–263
  ordering of, 277–279, 279e
  refining and testing, 273–274
Survey research
  attractions of, 255–257
  avoiding confusing phrasing in, 260–262
  comparison of designs for, 301–303, 302e
  defined, 255
  designing questionnaires for, 273–281
  errors in, 258–259
  ethical issues in, 303–304
  maintaining consistent focus in, 277
  omnibus, 258
  organizing, 281–303
  questionnaires made attractive, 280
  in the social sciences, 255–259
  translation of questionnaires in, 280–281
  writing questions for, 259–268
Survey Research Laboratory (SRL), 256, 287
Swanger, Rebecca Todd, 429
Swiss, Liam, 552, 567
Symbolic interaction theory, 34
Syme, S. Leonard, 277
Synchronous online interviews, 395–396
Systematic bias, 157
Systematic errors, 132
Systematic literature reviews, 44–46
Systematic observation, 446–448, 447–448e
Systematic random sampling, 162, 163e

Tabula rasa, 368
Tacit knowledge, 425–426
Take Charge of Your Life (TCYL), 511
Talk therapy, 493
Target population, 152
Tavory, Iddo, 430–431
Taylor, Bruce, 44–45
Teachers, preschool, 598
Tearoom Trade, 70, 82–83
Teenage Health in Schools, 114–115, 115e
Telephone surveys, 286–293, 289–292e
Temporal historical research, 554
Ten Point Coalition, 510
Testa, Maria, 465–466
    embedded designs, 471–472
Testing and causal validity, 239
Test-retest reliability, 136
Theoretical sampling, 385–386, 386e
Theory
    definition of, 33
    links between data and, 48e
    of moral development, 436
    program, 505–506, 507e
Theory-driven evaluation, 505
Therapeutic communities (TCs), 503–504
Thick description, 373
Thill, Kathryn, 575, 579, 582
Thoennes, Nancy, 58
Thomas, Kirsten, 84
Thompson, Sanna, 148, 169
Thorne, Barrie, 368, 378, 383, 384, 387, 389
    on reflexivity, 427
Thorsen, Maggie, 254–255
Time order, 193–195, 198, 206, 207e
Time series designs, 235–236
Timmermans, Stefan, 430–431
Tjaden, Patricia, 58
Toepoel, Vera, 596–597, 605
Topalli, Volkan, 402
Townsend, MEg, 495
Tradition, excessive devotion to, 10
Transitional Aid Research Project (TARP), 496, 498, 499e

Translation of surveys, 280–281
Trend studies, 196–197
Triangulation, 124–125, 467
Trow, Martin, 467
True experiments, 204, 224–231
    causality in, 230–231
    generalizability in, 242–245
    limitations of, 229–230
    See also Experiments/experimentation
Truman, Harry S., 158
Trump, Donald, 159, 327
Turkle, Sherry, 376, 418
Tuskegee Study of Untreated Syphilis in the Negro Male, 70–71

UCLA Center for Communication Policy, 19
Unbalanced response choices, 132, 133e
Uncritical agreement with authority, 10
Ungrouped data, 322–323e
Uniform Crime Reports (UCR), 122, 341, 560
Unimodal distributions, 328
Union Democracy, 467
United Nations, 524
Units of analysis, 189–192
    ecological fallacy and reductionism, 190–192, 191e
Units of analysis in content analysis, 578–579
Units of observation, 190
Univariate distributions
    displaying, 316–326
    frequency distributions, 322–326, 322–326e
    graphs, 318–321, 319–321e
    summarizing, 326–335
Unlabeled unipolar response options, 265, 266e
Unobtrusive measures, 120–121
    available data and, 121–122
    coding content and, 122
Unrepresentative samples, 154, 155e
Urban Institute, 600
USA Today, 579
Use and misuse of results, 457–458
Utility of response categories, 264–265
Utilization-focused evaluation, 508

Validity, 57–58
  causal, 58, 60–61, 238–242
  construct, 134–135
  content, 133
  criterion, 133–134, 135
  in experiments, 238–245
  external, 59, 244–245
  face, 132–133
  generalizability, 58, 59–60, 242–245
  internal, 58, 60–61, 238–242
  measurement, 58–59, 132–135
  ways to improve, 138–141
Values role in research, 20–21
Van Hoye, Greet, 243–244
Van Maanen, John, 379, 383, 403
Van Ryn, Michelle, 239
Van Winkle, Barrik, 453
VanZile-Tamsen, Carol, 465
Variability, 316
  distributions differing not in central tendency, but in, 331–332, 332e
Variable-oriented research, 561
Variables, 111
  concepts, indicators, and, 117e
  cross-tabulating (*See* Cross-tabulation)
  dependent, 51
  design coding procedures for, 579–583
  extraneous, 208, 347–348, 347–348e
  independent, 51
  intervening, 346–347, 346e
  specification, 348–350, 349–350e
  vs. constants, 112
Variance, 333, 334e
Variation, measures of, 331–335
Venkatesh, Sudhir, 402–403
Versatility of survey research, 255
Victory, Nancy, 13
Video ethnography, 446, 458
Vinokur, Amiram, 239
Violence Against Women, 465
Virtual community, 13
Virtual ethnography, 375
Visual sociology, 440–446

Visual Studies, 444
Viterna, Jocelyn, 552, 567
Volkow, Nora, 110
Voluntary participation, 401–402
Voting
    gender, interpersonal trust, and, 342, 342e
    household income and, 338–340, 339–340e, 346, 346e
    immigrants and, 561–563, 562–563e
    income and education and, 347–350, 348–350e
    international turnout, 561–565, 562e, 564–565e
    likelihood of, 313–314, 314–315e
Vulnerable populations, 153

Wageman, Ruth, 235
Walker, Henry A., 247
Ward, Lester Frank, 21
Ward Schofield, Janet, 176
Ware, Norma, 388
Warner, W. Lloyd, 466
Weaver-Hightower, Marcus, 468, 481
Webb, Eugene, 120–121
Weber, Max, 553, 567
Web surveys, 296–300, 296e, 299e
Wechsler, Henry, 104, 151
Weil, John, 28, 39
Weinberg, Darin, 116
Weisburd, David, 44–45
Well-being, subject, 402–403
Wellman, Barry, 16, 17
    applied research by, 20
    surveys used by, 19
Wells, William, 297
Wengraf, Tom, 393
Wertz, Frederick, 431
Westby, Ruth, 566
Western, Bruce, 30
WEXTOR, 243
Whelan, Christopher, 522
Whyte, William Foote, 379, 382, 383, 390, 449
    corroboration and legitimization of conclusions, 426
    subject well-being and, 402–403
Wikipedia, 612

Willer, David, 247
Williams, Kirk, 40
Wilson, William Julius, 611
Winiarek, Claire Wulf, 351
Winship, Chris, 542
Wittenbrink, Bernd, 120
Women's representation in politics, 552–553
Wood, Heather, 301
Work and Authority in Industry, 553
World Health Organization (WHO), 28, 281
World Income Inequality Database, 533
World War II, 490
Writing research, 592–594

Xu, Yili, 190

Yankelovich, Daniel, 509–510
Yavan, Tülay, 39
Yeater, Elizabeth, 414
Young, Michael, 611
Young adults
    binge drinking among, 103–105
    homelessness among, 148
    migration in China, 472–473
    sociopolitical attitudes and college attendance of, 312
    transition from adolescence to, 254–255
Youth gangs, 106–107
Youth & Society, 148
YouTube, 440

Zaslow, Martha, 505
Zelditch, Morris, Jr., 223–224
Zhong, Juan, 472–473, 479
Zimbardo prison simulation study/Philip Zimbardo, 70, 77–79, 83, 245–246